

Kokpity menedżerskie w analizie i prezentacji danych biznesowych z wykorzystaniem MS Excel 2016

pod redakcją
Doroty Buchnowskiej



UNIWERSYTET GDAŃSKI

Wydział Zarządzania
Uniwersytetu Gdańskiego
Sopot 2016

www.wzr.pl

Recenzent

prof. dr hab. Zygmunt Drązek

Projekt okładki i strony tytułowej

ESENCJA Sp. z o.o.

Redakcja

Jerzy Toczek

Skład i łamanie

Mariusz Szewczyk

© Copyright by Wydział Zarządzania Uniwersytetu Gdańskiego 2016

ISBN 978-83-64669-07-1

Wydawca

Wydział Zarządzania

Uniwersytetu Gdańskiego

81-824 Sopot, ul. Armii Krajowej 101

Druk

Zakład Poligrafii Uniwersytetu Gdańskiego,

Sopot, ul. Armii Krajowej 119/121

tel. 58-523-13-75, 58-523-14-49, e-mail: poligraf@gnu.univ.gda.pl

Spis treści

Wprowadzenie	7
Rozdział 1	
Pozyskiwanie danych na potrzeby analiz biznesowych	11
<i>Jacek Maślankowski</i>	
Wstęp	11
1.1. Rodzaje źródeł i formatów danych	11
1.2. Przypadek 1 – pozyskiwanie danych ze stron internetowych na temat importu i eksportu towarów	15
1.3. Przypadek 2 – pozyskiwanie danych z plików csv na temat sprzedaży towarów i usług	18
1.4. Przypadek 3 – pozyskiwanie danych z zewnętrznych źródeł danych na temat notowań kursów walut	19
1.5. Przypadek 4 – pozyskiwanie danych giełdowych	23
1.6. Przypadek 5 – pozyskiwanie danych z bazy Eurostat na temat wysokości wynagrodzeń i poziomu wykształcenia	26
1.7. Możliwe zastosowania MS Excel w biznesie w formie zadań	29
Zakończenie	29
Bibliografia	30
Rozdział 2	
Przygotowanie danych do analizy	31
<i>Dariusz Kralewski, Monika Woźniak</i>	
Wstęp	31
2.1. Przypadek 1 – przygotowanie danych na temat eksportu i importu z wykorzystaniem mechanizmów formatowania	32
2.2. Przypadek 2 – przygotowanie danych do analizy z wykorzystaniem formuł tekstowych	34
2.3. Przypadek 3 – wykorzystanie narzędzi danych w celu poprawy ich jakości	40
2.4. Możliwe zastosowania MS Excel w biznesie w formie zadań	43
Zakończenie	43
Bibliografia	44
Rozdział 3	
Analiza danych z wykorzystaniem zaawansowanych formuł i funkcji	45
<i>Michał Kuciapski</i>	
Wstęp	45
3.1. Przypadek 1 – analiza danych sprzedaży z wykorzystaniem funkcji matematycznych i statystycznych	46
3.2. Przypadek 2 – analiza danych sprzedaży z zastosowaniem funkcji logicznych	50

3.3	Przypadek 3 – analiza danych struktury wykształcenia w oparciu o funkcje wyszukiwania i adresu	52
3.4	Przypadek 4 – analiza danych sprzedaży z wykorzystaniem funkcji bazodanowych	54
3.5	Przypadek 5 – analiza danych sprzedaży z wykorzystaniem funkcji finansowych	57
3.6.	Możliwe zastosowania MS Excel w biznesie w formie zadań	59
	Zakończenie	60
	Bibliografia	61
	Rozdział 4	
	Graficzna prezentacja danych biznesowych.	63
	<i>Bartosz Marcinkowski</i>	
	Wstęp	63
4.1.	Przypadek 1 – wykorzystanie formatowania warunkowego na potrzeby analizy wynagrodzeń	64
4.2.	Przypadek 2 – wykresy przebiegu w czasie jako alternatywna forma analizy dynamiki zjawiska w czasie	67
4.3.	Przypadek 3 – graficzna prezentacja kursów walut w oparciu o różne odmiany wykresów tradycyjnych	68
4.4.	Przypadek 4 – wizualizacja sprzedaży z wykorzystaniem wykresów specjalistycznych	72
4.5.	Przypadek 5 – analiza porównawcza odsetka osób z wyższym wykształceniem z wykorzystaniem map	75
4.6.	Możliwe zastosowania MS Excel w biznesie w formie zadań	79
	Zakończenie	79
	Bibliografia	80
	Rozdział 5	
	Analiza danych z wykorzystaniem mechanizmów filtrowania i tabel przestawnych	83
	<i>Dorota Buchnowska</i>	
	Wstęp	83
5.1.	Przypadek 1 – wykorzystanie sum częściowych w analizie sprzedaży	84
5.2.	Przypadek 2 – zastosowanie mechanizmów filtrowania w celu dostosowania danych do potrzeb informacyjnych decydentów	88
5.3	Przypadek 3 – filtrowanie danych z zastosowaniem formuł w analizie importu i eksportu towarów	91
5.4.	Przypadek 4 – tabele przestawne w analizie sprzedaży produktów	93
5.5.	Przypadek 5 – wykorzystanie fragmentatorów i osi czasu do filtrowania oraz integracji tabel i wykresów przestawnych	98
5.6.	Możliwe zastosowania arkusza MS Excel w formie zadań	102
	Zakończenie	102
	Bibliografia	103

Rozdział 6

Tworzenie modelu danych z wykorzystaniem Power Pivot 105

Dariusz Kralewski

Wstęp 105

6.1. Przypadek 1 – integracja kursów walut, danych giełdowych oraz danych
sprzedaży 107

6.2. Przypadek 2 – integracja danych dotyczących sprzedaży 111

6.3. Przypadek 3 – wykorzystanie języka DAX w analizie danych na temat
sprzedaży 1126.4. Przypadek 4 – integracja danych dotyczących importu i eksportu
owoców 114

6.5. Możliwe zastosowania MS Excel w biznesie w formie zadań 115

Zakończenie 116

Bibliografia 117

Rozdział 7

Podstawy tworzenia kokpitu menedżerskiego z wykorzystaniem Power Pivot 119

Jacek Maślankowski

Wstęp 119

7.1. Przypadek 1 – interaktywny kokpit do analizy notowań giełdowych
i kursów walut 1207.2. Przypadek 2 – kokpit menedżerski z wykorzystaniem KPI do analizy
i oceny sprzedaży 1257.3. Przypadek 3 – integracja wielu źródeł danych w jednym kokpicie
menedżerskim 129

7.4. Możliwe zastosowania MS Excel w biznesie w formie zadań 131

Zakończenie 131

Bibliografia 132

Wykorzystane skróty 133

Pliki z danymi oraz wynikami analizy danych 135

Nota o autorach 137

Wprowadzenie

Skuteczne i efektywne zarządzanie organizacją gospodarczą wymaga podejmowania decyzji na podstawie wiedzy bazującej na dobrej jakości danych i informacjach. Rozwój oraz upowszechnienie technologii informatycznych sprawiły, że firmy mają szeroki dostęp do danych zarówno w skali mikro – na poziomie organizacji – jak również w skali makro. Rosnące zasoby informacji sprawiają jednak, że proces podejmowania decyzji jest coraz trudniejszy i wymaga prowadzenia zaawansowanej analizy danych w wykorzystaniem rozwiązań IT. Analityka biznesowa rozumiana jako wydobywanie użytecznych informacji i wiedzy z dużych wolumenów danych w celu doskonalenia procesu podejmowania decyzji biznesowych jest dla współczesnego przedsiębiorstwa determinantą wysokiej pozycji firmy na konkurencyjnym rynku. Analizie dużej ilości danych dedykowane są rozwiązania klasy Business Intelligence i Big Data, jednak ze względu na poziom zaawansowania i koszty wykorzystywane są one przede wszystkim przez duże przedsiębiorstwa. Rozwiązaniem analityki biznesowej, które może być skutecznie wykorzystywane, szczególnie przez mniejsze organizacje, jest arkusz kalkulacyjny MS Excel.

Celem niniejszej publikacji jest zaprezentowanie możliwości MS Excel jako narzędzia kompleksowo wspierającego proces analizy danych na potrzeby podejmowania decyzji biznesowych. W siedmiu rozdziałach pokazano kolejne etapy przetwarzania różnego rodzaju danych – począwszy od ich pozyskania, a skończywszy na ich spersonalizowanej prezentacji w postaci kokpitu menedżerskiego – z wykorzystaniem najnowszej wersji MS Excel 2016 (premiera we wrześniu 2015). Wbudowane w to rozwiązanie narzędzia, takie jak Power Pivot i Power View, ze względu na zaawansowane możliwości analizy i prezentacji danych określane są mianem Power BI.

Pierwszy rozdział omawia problematykę związaną z pozyskiwaniem i pobieraniem do MS Excel danych z różnych źródeł. Zaprezentowano w nim mechanizmy importowania do MS Excel danych, zarówno z wewnętrznych systemów transakcyjnych, jak i wiarygodnych zewnętrznych źródeł, takich jak baza Eurostat. Dzięki szerokiemu przekrojowi danych firma może nie tylko monitorować i ewaluować wewnętrzne procesy w przedsiębiorstwie, ale również sprawnie i prawidłowo reagować na wszelkie zmiany w otoczeniu.

Pozyskane z różnych źródeł dane często zawierają liczne błędy i są niespójne. W drugim rozdziale ukazano mechanizmy poprawy jakości danych dostępne w MS Excel. Zastosowanie zaprezentowanych metod oraz reguł przetwarzania danych jest niezbędnym etapem przygotowania danych do dalszych analiz,

w znacznym stopniu wpływającym na wysoką jakość uzyskanej na ich podstawie informacji.

Trzeci rozdział opisuje możliwości przetwarzania odpowiednio przygotowanych danych z wykorzystaniem zaawansowanych formuł i funkcji w celu uzyskania podstawowych wskaźników ekonomicznych. Pozwalają one dokonywać ewaluacji bieżącej sytuacji przedsiębiorstwa w różnych obszarach i wymiarach oraz oceniać perspektywy rozwoju organizacji w kontekście panujących warunków rynkowych.

Równie ważna jak sama analiza jest właściwa prezentacja danych, która wpływa na poprawną interpretację uzyskanych informacji i w znacznym stopniu przyspiesza proces podejmowania decyzji. Zarówno zakres, jak i forma muszą być dostosowane do rodzaju danych i potrzeb decydenta. Dlatego też w rozdziale czwartym zaprezentowano najważniejsze i najnowocześniejsze metody wizualizacji (np. mapy tematyczne, wykresy w formie drzewa czy też wykresy wodospadowe) danych biznesowych, zwracając szczególną uwagę na reguły ich prawidłowego stosowania.

Większość danych w przedsiębiorstwie pochodzi z ewidencyjno-operacyjnych systemów, takich jak ERP, CRM czy SCM. Dane pochodzące z baz transakcyjnych to jednak zazwyczaj dane szczegółowe, np. pojedyncze transakcje sprzedaży, gdy tymczasem menedżer potrzebuje danych zagregowanych, przekrojowych, takich jak całkowity zysk ze sprzedaży konkretnej grupy produktów w wybranym okresie. Dlatego też przy analizie dużych zbiorów danych szczegółowych wykorzystywane są tabele i wykresy przestawne, które pozwalają na szybkie agregowanie i prezentowanie danych w różnych układach. Możliwości interaktywnej analizy danych z wykorzystaniem tabel i wykresów przestawnych oraz mechanizmów filtrowania danych zaprezentowane zostały w kolejnym, piątym rozdziale publikacji.

Posiadanie kompleksowych danych wiąże się z koniecznością opierania się na wielu różnych, często niepowiązanych ze sobą źródłach. Aby uzyskać holistyczny obraz analizowanej sytuacji, konieczne jest wówczas opracowanie zintegrowanego modelu danych. W rozdziale szóstym pokazano, jakie możliwości w tym zakresie daje rozwiązanie Power Pivot, które w kolejnym rozdziale wykorzystywane jest do tworzenia podstawowego narzędzia pracy każdego decydenta – kokpitu menedżerskiego.

Rozdział siódmy prezentuje teoretyczne i praktyczne aspekty tworzenia kokpitów menedżerskich, których głównym zadaniem jest przekazywanie relewantnej informacji decydującym, prezentowanej w syntetycznym i analitycznym ujęciu. Zakres wykorzystania kokpitów w biznesie jest coraz szerszy, zarówno pod względem obszaru zastosowań, jak i poziomu zarządzania. Jednocześnie są to narzędzia mocno personalizowane. Z tego też względu budowa kokpitów wymaga

zastosowania mechanizmów integracji danych pochodzących z różnych źródeł oraz wielu różnorodnych narzędzi analitycznych, które omówiono we wcześniejszych rozdziałach. Jednym z ważnych zagadnień poruszanych w rozdziale siódmym jest również budowa wskaźników KPI, które są nieodłącznym elementem każdego kokpitu menedżerskiego.

Dla pełniejszego zobrazowania możliwości i ograniczeń omawianych w niniejszej publikacji narzędzi ich zastosowanie ukazano na przykładzie konkretnej firmy, wykorzystując je do rozwiązywania wybranych problemów biznesowych. W każdym z rozdziałów zawarte zostały studia przypadków, które kolejno pokazują, jak powinien przebiegać proces pozyskania, a następnie przekształcenia surowych danych w użyteczną wiedzę służącą podejmowaniu właściwych decyzji w zakresie zarządzania przedsiębiorstwem z wykorzystaniem narzędzia MS Excel.

Dorota Buchnowska

Rozdział 1

Pozyskiwanie danych na potrzeby analiz biznesowych

Jacek Maślankowski

Wstęp

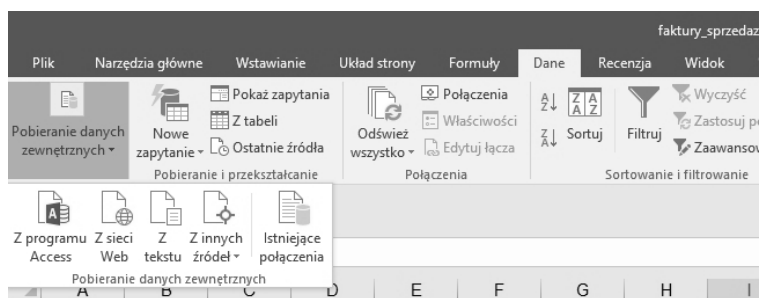
Od wielu lat podkreśla się znaczenie danych dla firmy, stanowiących jedną z podstawowych wartości dla przedsiębiorstwa. Bardzo często podkreśla się w literaturze biznesową wartość informacji. Kluczem jednak jest właściwe zaplanowanie i wdrożenie systemu zarządzania wiedzą, czyli przetworzonymi danymi [Mousavizadeh i in., 2015]. Przetwarzane w firmie dane najczęściej pochodzą ze źródeł wewnętrznych, czyli znajdujących się w firmie. Przykładem są chociażby dane na temat wielkości sprzedaży pozyskiwane na podstawie faktur. Aby możliwe było zbadanie otaczającego środowiska, należy również sięgnąć do danych zewnętrznych, które pochodzą najczęściej z publicznie dostępnych źródeł lub są kupowane w firmach zajmujących się dystrybucją różnego rodzaju danych. Dane publicznie dostępne mogą dotyczyć koniunktury gospodarczej czy też wielkości importu i eksportu różnego rodzaju towarów. Dostęp do takich danych pozwoli firmie zaplanować rozwój i ewentualną ekspansję na rynki poza granice kraju. Jednocześnie możliwe będzie usprawnienie procesu podejmowania decyzji, przede wszystkim na szczeblu strategicznym, dzięki dostępowi do szerszego spektrum analiz danych. Istotne jest zatem zapewnienie decydom dostępu do tego rodzaju informacji. W niniejszym rozdziale skupiono się na problematyce związanej z dostępem do szeroko rozumianych danych zewnętrznych. Opisane w kolejnych rozdziałach studia przypadków pozwalają na poznanie metod oraz reguł przetwarzania i analiz danych w firmach z zastosowaniem powszechnie występujących rozwiązań w arkuszach kalkulacyjnych. W stosowanych studiach przypadków wykorzystano arkusz Microsoft Excel jako najpopularniejszy przykład arkusza kalkulacyjnego stosowanego powszechnie w firmach.

1.1. Rodzaje źródeł i formatów danych

Wyróżnia się wiele formatów, które mogą stanowić źródło danych dla arkusza kalkulacyjnego MS Excel. Najpopularniejszym źródłem są dane wpisywane bez-

pośrednio w arkuszu lub wczytywane poprzez otwarcie pliku ustrukturyzowanego z rozszerzeniem xls/xlsx lub częściowo ustrukturyzowanego posiadającego rozszerzenie csv. Ustrukturyzowanie danych dotyczyć może nie tylko podziału na kolumny i wiersze, ale również informacji o zastosowanych w zbiorze typach danych. Dane takie mogą później zostać poddane dalszej obróbce i wykorzystane do analiz przez szeroko pojęte systemy klasy Business Intelligence [Wrycza, 2010].

W niniejszym podrozdziale skoncentrowano się na metodzie dostępu do danych z wykorzystaniem arkusza kalkulacyjnego MS Excel w wersji 2016. Dostęp do zewnętrznych zbiorów danych uzyskuje się poprzez wykorzystanie wstążki *Dane*, pokazanej na rysunku 1.1.



Rysunek 1.1. Narzędzia do pobierania danych w arkuszu kalkulacyjnym MS Excel

Źródło: Opracowanie własne.

Przedstawione na rysunku 1.1 narzędzia pozwalają odpowiednio na dostęp do danych pobieranych z bazy MS Access (pliki z rozszerzeniem mdb lub accdb), z sieci Web (zasoby dostępne na stronach internetowych), z tekstu (pliki tekstowe częściowo ustrukturyzowane) oraz z innych źródeł, widocznych na rysunku 1.2 i omówionych poniżej. Istnieje również możliwość dostępu do danych w ramach wcześniej skonfigurowanych połączeń.

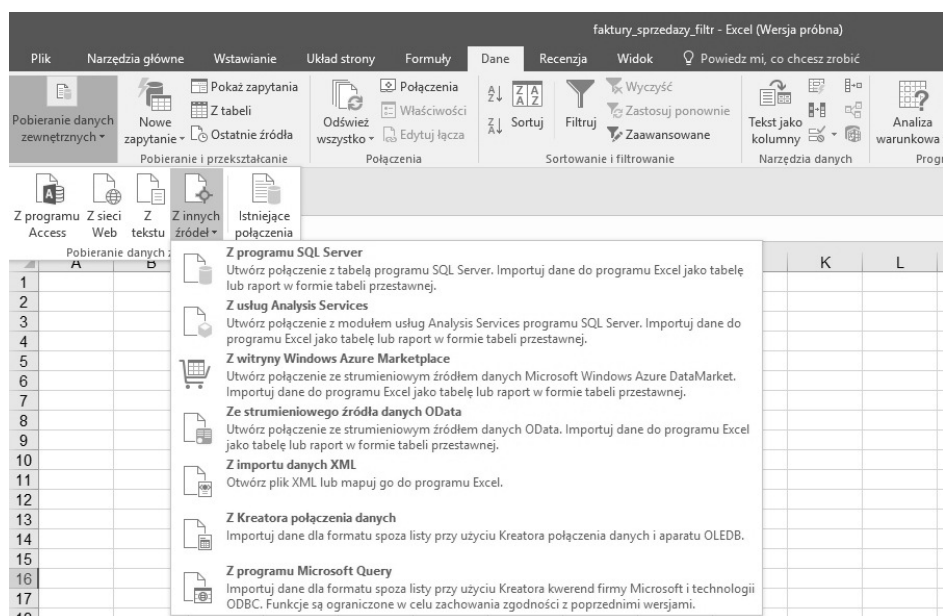
Dostęp do danych z bazy MS Access wymaga posiadania tej bazy danych w postaci plikowej. Oznacza to, że nie jest wymagane narzędzie MS Access, aby mieć dostęp do tego rodzaju baz danych. Narzędzie MS Excel posiada wbudowany interfejs dostępowy, umożliwiający podłączenie się do pliku bazodanowego MS Access.

Bardziej złożony jest przykład z dostępem do zasobów w sieci Web. Zasoby te, ze względu na różnorodność formatu ich zapisu, nie zawsze pozwalają na poprawne wczytanie do arkusza MS Excel. W takim wypadku niezbędne jest oczyszczenie danych, którego celem jest usunięcie wszelkiego rodzaju problemów związanych z błędami w danych, pojawiających się podczas ich importowania. Do takich błędów zalicza się na przykład przesunięcie w kolumnie, gdy

dwuczłonowe nazwisko zostało wpisane w dwóch kolejnych kolumnach zamiast w jednej. Tego rodzaju błędy poprawiane są w kolejnym etapie zasilania systemów informatycznych danymi, zwanym wstępnym przetwarzaniem danych. Szerzej zostało to zaprezentowane w rozdziale drugim niniejszego opracowania.

Podobne problemy jak z importowaniem z sieci Web występują w przypadku importowania danych z pliku tekstowego. Tego rodzaju dane są w szczególności obciążone błędami niespójności w zapisie danych. Bardzo często dane takie są importowane jako tekst. Aby móc wykorzystywać potencjał narzędzi analitycznych, należy wcześniej te dane przekształcić na format liczbowy.

Bardziej zaawansowane zastosowanie arkusza kalkulacyjnego ma miejsce podczas podłączania się do zewnętrznych źródeł danych w celu ich zaimportowania. Zobrazowano to na rysunku 1.2.



Rysunek 1.2. Pobieranie danych z zewnętrznych źródeł w arkuszu kalkulacyjnym MS Excel
Źródło: Opracowanie własne.

Widoczne na rysunku 1.2 źródła danych dotyczą pobierania danych z zewnętrznej bazy relacyjno-obiektowej MS SQL Server (ustrukturyzowane tabele posiadające wzajemne relacje) oraz z usług *Analysis Services* (kostki wielowymiarowe OLAP). Są to najczęściej występujące w biznesie odwołania do baz danych, które posiadają z góry określoną strukturę danych [Tripathy, Das, 2011]. Dodatkowo MS Excel umożliwia dostęp do ogólnie dostępnych danych z witryny MS Azure Marketplace (usługa dostępna w tzw. chmurze obliczeniowej), strumieniowego źródła danych Odata, plików XML, kreatora połączeń danych oraz progra-

mu Microsoft Query do tworzenia kwerend w postaci kreatora dla źródeł ODBC (ang. *Open Database Connectivity*). ODBC umożliwia niemal nieograniczony dostęp do wszelkiego rodzaju baz danych, takich jak Oracle czy Teradata. Zasady działania tego mechanizmu zostały szczegółowo opisane w [Wood, 1999].

Przykładami źródeł danych, które mogą zostać podłączone do narzędzia MS Excel poprzez wymienione powyżej interfejsy, są bazy danych systemów CRM (wspierające zarządzanie relacjami z klientami), ERP (planowanie zasobów przedsiębiorstwa) czy też SCM (zarządzanie łańcuchem dostaw). Bazy te najczęściej występują w modelu relacyjnym, co sprawia, że dostęp do nich jest możliwy poprzez SQL Server oraz ODBC. W niniejszym rozdziale poruszono również kwestie dostępu do wielowymiarowych kostek analitycznych OLAP. Ich istotę oraz mechanizm działania opisano szerzej w książce [Wrycza, 2010, s. 421–423].

Z powyższego tekstu jednoznacznie wynika, że analityk biznesowy ma bardzo szerokie możliwości dostępu do danych biznesowych zapisanych w większości popularnych formatów. Istotną kwestią, jaką należy wziąć pod uwagę, jest forma ustrukturyzowania danych. Informacja przechowywana w postaci kolumn i wierszy nie ma charakteru ściśle ustrukturyzowanego. Takie źródła określa się jako częściowo ustrukturyzowane.

W pełni ustrukturyzowane źródła danych to takie, które posiadają nie tylko precyzyjnie zdefiniowane kolumny i wiersze, ale również typy danych, jakie się znajdują w źródle. Typy danych dostępne w MS Excel zostały wymienione w tabeli 1.1.

Tabela 1.1. Najważniejsze typy danych w arkuszu kalkulacyjnym MS Excel

Typ danych MS Excel	Opis
Walutowy (ang. <i>currency</i>)	Zwykle określa walutę lokalną. Dla polskiej opcji lokalizacyjnej będzie to PLN.
Daty i czasu (ang. <i>datetime</i>)	Służy do przechowywania daty i czasu w celu prowadzenia obliczeń, np. liczby dni od danego.
Logiczne (ang. <i>logical</i>)	Są przydatne przy definiowaniu wyrażeń logicznych, np. instrukcji warunkowych. Przechowują wartości prawda lub fałsz.
Liczbowe (ang. <i>number</i>)	Dzieli się na całkowite i zmiennoprzecinkowe. Zmiennoprzecinkowe pozwalają na zapisywanie wartości ułamkowych.
Tekstowe (ang. <i>text</i>)	Służy do przechowywania tekstu.

Źródło: Opracowanie własne na podstawie [msdn.microsoft.com, dostęp dnia 19.07.2015].

Podczas importowania danych biznesowych może się okazać, że MS Excel potraktował dane liczbowe jako tekst, poprzedzając liczby w komórkach apostrofami ('). Nie jest wówczas możliwe liczenie sum czy wykonywanie jakichkolwiek formuł bazujących na liczbach, np. statystycznych lub prostych matematycznych.

O sposobie konwersji takich danych można przeczytać w rozdziale 2 niniejszego opracowania.

Dane pozyskiwane na potrzeby biznesu mogą mieć również charakter nieustrukturyzowany, tj. najczęściej zapisane są postaci tekstu i wymagają analizy za pomocą złożonych metod. Tego rodzaju dane powinny być przetwarzane przez wyspecjalizowane systemy, np. Big Data [Kuiler, 2014] lub z wykorzystaniem metod *text mining*, czyli znajdowania wzorców w tekście w celu pozyskiwania wartościowej informacji [Kim i in., 2014]. Takie dane przed ich oczyszczeniem zwykle gromadzi się w bazach danych typu NoSQL, które nie mają typowo tabelarycznej struktury, lecz najczęściej przechowują dane na zasadzie parowania klucz–wartość [Duda, 2012].

Celem niniejszego rozdziału jest przegląd możliwości dostępu do źródeł danych biznesowych dla firm i zaproponowanie rozwiązań pozwalających na ich sprawną analizę z wykorzystaniem arkusza kalkulacyjnego.

1.2. Przypadek 1 – pozyskiwanie danych ze stron internetowych na temat importu i eksportu towarów

W przypadku firm prowadzących handel międzynarodowy kluczowe są zagadnienia związane z handlem zagranicznym. Takie dane można pozyskać m.in. z baz danych zawierających dane statystyczne na temat importu i eksportu towarów i usług. W ramach opisywanego studium przypadku przyjęto założenie, że do realizacji analiz niezbędny jest zbiór danych pochodzący ze strony internetowej. Zbiory takie często są generowane na podstawie znajdujących się w Internecie baz danych. Do takich baz danych zaliczyć można międzynarodowe bazy, które gromadzą dane dotyczące wielu krajów. Dane te jednocześnie można porównywać ze względu na jednolitą formę zapisu.

Źródłem danych zastosowanym w niniejszym studium przypadku dla Firmy X, która zajmuje się handlem owocami i warzywami, jest baza Hinx udostępniana na stronach Głównego Urzędu Statystycznego. Dostęp do tej bazy jest możliwy bezpośrednio przez wejście na stronę internetową o adresie <http://www.stat.gov.pl>, a następnie wybranie z grupy *Banki i bazy danych* pozycji *Handel zagraniczny*. Baza ta zawiera podstawowe dane na temat importu oraz eksportu towarów i usług. Dostęp do bazy danych jest możliwy poprzez interfejs WWW. Wygenerowane dane są prezentowane bezpośrednio w przeglądarce internetowej. Istnieje jednak możliwość ich wyeksportowania do jednego z najpopularniejszych formatów plików z rozszerzeniem doc (MS Word), xls (MS Excel) lub pdf (Adobe Acrobat). Przykład udostępnianych formatów plików znajduje się na rysunku 1.3.

Przeglądanie danych
Przeglądanie, wyszukiwanie obiektów zarejestrowanych w systemie

kraj	kod CN	nazwa CN	ilość	masa	jednostka miary	wart. w złotych	wart. w dolarach	wart. w euro
Bangladesz	880220	Samoloty i pozostałe statki powietrzne, bez śmigłowców, o masie własnej <= 2000 kg	1			1009038	270100	233476
Chiny	880220	Samoloty i pozostałe statki powietrzne, bez śmigłowców, o masie własnej <= 2000 kg	3		szt.	1118619	313902	259782
Francja	880220	Samoloty i pozostałe statki powietrzne, bez śmigłowców, o masie własnej <= 2000 kg	1		szt.	251638	68410	60050
Niemcy	880220	Samoloty i pozostałe statki powietrzne, bez śmigłowców, o masie własnej <= 2000 kg	1		szt.	216504	61350	50280
Szwajcaria	880220	Samoloty i pozostałe statki powietrzne, bez śmigłowców, o masie własnej <= 2000 kg	2		szt.	690	185	165
Zjednoczone Emiraty Arabskie	880220	Samoloty i pozostałe statki powietrzne, bez śmigłowców, o masie własnej <= 2000 kg OGÓLEM	2		szt.	59773	16000	13831

liczba rekordów 7 rekordów na stronie 10 numer strony 1

eksport do pliku PDF (.pdf) Word (.doc) Excel (.xls) PDF (.pdf)

Rysunek 1.3. Opcje eksportowania bazy danych handlu zagranicznego do zewnętrznych plików

Źródło: Opracowanie własne.

Aby dojść do sytuacji zaprezentowanej na rysunku 1.3, należy wybrać z bazy danych Hinex opcję *Eksport*, następnie rok, za który należy pozyskać dane, oraz symbol według nomenklatury scalonej (ang. CN – *combined nomenclature*). Pełna klasyfikacja produktów i usług według nomenklatury scalonej jest dostępna na stronie Eurostatu, <http://ec.europa.eu/eurostat/ramon/nomenclatures/> pod nazwą *Combined Nomenclature*. W zależności od roku referencyjnego klasyfikacja ta może posiadać drobne zmiany w produktach. Wersja polska nomenklatury scalonej publikowana jest corocznie na stronach Głównego Urzędu Statystycznego <http://stat.gov.pl/sprawozdawczosc/intrastat/nomenklatura-scalona/>.

Na potrzeby realizacji niniejszego przykładu pozyskano tabelę z bazy Hinex, składającą się z następujących kolumn:

1. Kod CN – nomenklatura scalona umożliwiająca jednoznaczne rozróżnienie oferowanego towaru lub grupy towarów.
2. Nazwa towaru – nazwa przyporządkowana do kodu CN.
3. Kraj – nazwa przyporządkowana do symbolu kraju.
4. Symbol kraju – dwuliterowy symbol definiujący kraj zgodnie z normą ISO-3166.
5. Jednostka miary – zastosowana jednostka miary.
6. Ilość – wyeksportowana lub zaimportowana ilość towaru wyrażona w jednostce miary.
7. Wartość w zł – wartość eksportu lub importu w zł.
8. Wartość w euro – wartość eksportu lub importu wyrażona w euro przeliczonym po średnim kursie w danym roku.
9. Wartość w dolarach – wartość eksportu lub importu wyrażona w euro przeliczonym po średnim kursie w danym roku.

Zbiór ten (plik *r1_import_i_eksport_no.xlsx*) znajduje się w czterech osobnych arkuszach, odpowiednio obrazujących eksport oraz import w 2013 i 2014 roku. Dane takie przed dalszą obróbką należy oczyścić, tj. usunąć wszelkie zbędne wiersze (wartości ogółem, sumy częściowe) i usunąć wszelkie scalone komórki, które występują w wybranych wierszach. Na rysunku 1.4 pogrubioną czcionką zaznaczono wiersze zawierające dane zagregowane, które uniemożliwiają przetwarzanie danych jako zbioru wielu zmiennych agregowalnych. Innymi słowy, sumowanie wartości ze wszystkich wierszy w danej kolumnie zwróci wynik zawyżony, ze względu na zawarcie w wierszach *RAZEM POZYCJA: (1)* oraz *RAZEM Kraje Eur. Śr.-Wsch. (2)* i *RAZEM Kraje Rozwijające się: (3)* wartości dla wszystkich krajów, które również znajdują się w tabeli.

080810	Jabłka świeże	RAZEM POZYCJA: (1)	kg	1 216 293 773	1 824 610 677	439 125 345	581 170 519	
080810	Jabłka świeże	RAZEM Kraje Eur. Śr.-Wsch. : (2)	kg	874 796 796	1 323 688 789	318 835 059	421 566 568	
080810	Jabłka świeże	Białoruś	BY	kg	144 932 878	188 966 719	45 487 274	60 250 881
080810	Jabłka świeże	Moldawia	MD	kg	138 500	189 401	45 422	59 421
080810	Jabłka świeże	Rosja	RU	kg	676 316 272	1 062 010 918	255 816 513	338 277 440
080810	Jabłka świeże	Ukraina	UA	kg	53 409 146	72 521 751	17 485 850	22 978 826
080810	Jabłka świeże	RAZEM Kraje rozwijające się : (3)	kg	63 174 939	90 760 656	21 846 943	28 946 745	
080810	Jabłka świeże	Algieria	DZ	kg	3 604 259	7 547 080	1 813 110	2 402 168
080810	Jabłka świeże	Antarktyda	AQ	kg	110	1 125	265	355
080810	Jabłka świeże	Arabia Saudyjska	SA	kg	37 191	97 407	23 430	30 864
080810	Jabłka świeże	Azerbejdżan	AZ	kg	369 040	527 857	127 011	169 424
080810	Jabłka świeże	Bahrajn	BH	kg	18 922	72 479	17 427	23 234
080810	Jabłka świeże	Bośnia i Hercegowina	BA	kg	5 291 565	4 862 455	1 175 499	1 552 392
080810	Jabłka świeże	Czarnogóra	ME	kg	141 960	158 711	38 040	49 494
080810	Jabłka świeże	Egipt	EG	kg	1 544 918	3 094 810	746 336	998 175

Rysunek 1.4. Zbiór danych dotyczących handlu zagranicznego

Źródło: Opracowanie własne.

Podsumowując, zaimportowany zbiór danych zawiera usterki, które powinny zostać usunięte przed dalszą obróbką. Do takich usterek w istniejącym zbiorze zaliczyć należy:

- scalone komórki,
- przesunięcia pomiędzy kolumnami arkusza,
- wiersze zawierające sumy częściowe.

Sposób rozwiązania takich problemów został przedstawiony w rozdziale 2. Pozostawienie nieoczyszczonych danych może doprowadzić do utraty jakości danych definiowanej przez dwa podstawowe wymiary: kompletność i spójność [Kwon i in., 2014]. Kompletność w tym przypadku oznacza niewystępowanie braków danych. Natomiast spójność oznacza możliwość porównywania danych.

1.3. Przypadek 2 – pozyskiwanie danych z plików csv na temat sprzedaży towarów i usług

Firma X prowadzi sprzedaż i dystrybucję produktów spożywczych – warzyw oraz owoców. Dane transakcyjne na temat bieżących obrotów zawarte są w systemach informatycznych, m.in. ERP i CRM, które zapisują je w różnego rodzaju formatach, w tym również tekstowych wykorzystujących różne strony kodowe znaków.

Przygotowany zbiór danych (plik *r1_dane_sprzedaży_no.csv*) składa się z danych dotyczących sprzedaży produktów spożywczych. Plik ten zawiera dużą liczbę błędów edytorskich, które wymagają usunięcia przed dalszą obróbką danych. Błędy te można zauważyć na rysunku 1.5.

Biał,a Poc	Białystok	Bielsko-Bi	Bydgoszcz	Białystok	Bytom (ŁS C
29138.223	11705.036	4764.3450	18776.967	22763.370	20824.802 2
13015.539	1610.2216	17985.411	22809.003	8858.1567	5725.2919 2
27731.269	23881.945	4566.9432	27918.261	16198.859	11454.018 6
4018.4312	19010.332	28909.999	20145.917	16770.540	22534.956 1
29727.190	15511.993	19928.501	22272.998	16872.102	27912.895 1
7832.2543	29400.009	14562.135	1484.5611	9256.1644	6468.6349 2
18851.897	17603.754	26696.776	11497.434	5008.5814	3928.1308 2
16705.795	17531.110	27792.096	18998.903	5099.7331	23828.482 2
3969.1373	27499.058	1658.1924	3116.9737	29211.261	20940.873 2
3142.4785	18080.834	9829.1380	23065.892	20635.915	17050.950 2

Rysunek 1.5. Zbiór danych z błędnie przypisaną stroną kodową znaków

Źródło: Opracowanie własne.

Przykładowe zagadnienia, jakie analityk danych musi rozpatrzyć w takim zbiorze, obejmują:

- w przypadku gdy w danych występują klasyfikacje, należy przyporządkować im kody (np. dla jabłek przypisać kod 1, dla gruszek kod 2 itd.);
- w niektórych przypadkach konieczne jest wyliczenie wartości, jeżeli zbiór posiada wartość ogółem, a dodanie poszczególnych pozycji na liście nie równa się tej wartości, wówczas najczęściej należy dodać pozycję *inne* lub *pozostałe*, będącą różnicą pomiędzy wartością ogółem a pozostałymi wartościami;
- należy zwracać uwagę na sposób zapisu liczb – niekiedy kodowanie jest niejednolite – np. cyfry dziesiętne po kropce, po przecinku itd.;
- w tradycyjnych klasyfikacjach należy zastosować hierarchię – np. jeżeli są dane na województwa, należy dodać agregat Polska; jeżeli na powiaty lub miasta – agregat Województwa i Polska;

- jeżeli dane są w różnych jednostkami miar – należy je w miarę możliwości ujednoczyć;
- strony kodowe mogą być niezgodne z systemem – należy wówczas na etapie otwierania zbioru danych wybrać właściwą stronę kodową;
- dane mogą być niejednolite np. dolnośląskie, dolnoslaskie – należy te dane oczyścić przed rozpoczęciem przetwarzania;
- z kolumn należy wybrać jak najwięcej atrybutów, np. podzielić kolumnę z datą na wiele kolumn: rok, miesiąc, dzień, słownie itd.

Podsumowując, analityk przed analizą zbioru danych jest zobowiązany do przygotowania tego zbioru tak, aby spełniał stawiane w danej firmie wymogi jakościowe.

1.4. Przypadek 3 – pozyskiwanie danych z zewnętrznych źródeł danych na temat notowań kursów walut

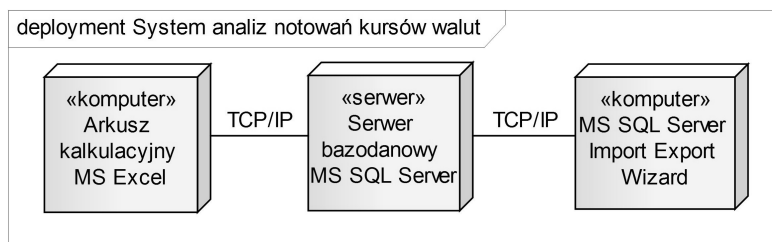
Bardzo często zdarza się, że dane mające zostać poddane analizie pochodzą z bazy lub hurtowni danych. Źródła te odpowiednio będą stanowiły powiązany zestaw tabel w relacyjnym modelu danych. Dla firmy, która prowadzi handel międzynarodowy, istotnym czynnikiem warunkującym jej rozwój jest kurs waluty kraju, z którym prowadzona jest wymiana handlowa. Informacja taka jest niezbędna również dla tych podmiotów, które kupują półprodukty z różnych regionów świata. Zakup takich materiałów, gdy złotówka się umacnia, może spowodować znaczące obniżenie kosztów produkcji.

W niniejszym podrozdziale rozważono przypadek Firmy X, która planowała eksportować oferowane produkty do krajów Unii Europejskiej. Istotne było, aby rozliczenia realizowano w czasie, gdy polska waluta umacniała się w stosunku do innych walut. Było to bardzo korzystne dla Firmy X, gdyż rachunki tej firmy prowadzone były w polskich złotych, a wymiana waluty następowała w momencie rozliczenia transakcji. Jednocześnie firma chciała archiwizować wszystkie gromadzone dane.

Firma X do przeprowadzenia analiz walut wykorzystywała zasoby ogólnodostępne w Internecie. Ze względu na duże rozmiary tabel archiwalnych z walutami, najlepszym sposobem przygotowania zbioru danych do analiz okazało się jego wczytanie do relacyjnej bazy danych MS SQL Server, a następnie dostęp do tego zbioru w ramach narzędzia MS Excel. Pozwala to równocześnie na archiwizację tych danych, gdyż z poziomu MS SQL Server administrator może zautomatyzować m.in. sporządzanie kopii zapasowych.

W tym celu w Firmie X zdecydowano się na wykorzystanie narzędzia MS Import Export Wizard, które jest ogólnie dostępnym oprogramowaniem w ramach

aplikacji klienckiej MS SQL Server. Ogólny schemat działania tego rozwiązania został zaprezentowany na diagramie wdrożeniowym na rysunku 1.6.



Rysunek 1.6. Proponowane wdrożenie systemu analiz kursów walut dla Firmy X

Źródło: Opracowanie własne.

Przedstawiony na rysunku 1.6 schemat obejmuje serwer bazodanowy umiejscowiony w centralnym miejscu. Dane są ładowane do serwera bazodanowego przez administratora systemu. Istotne jest, aby codziennie dane były aktualizowane. Analityk danych wykorzystuje arkusz kalkulacyjny MS Excel, aby pobierać dane z serwera bazodanowego.

Bieżące dane kursu walut miały być pobierane ze strony: <http://www.nbp.pl>. Aby skorzystać z bazy danych walut, należy z menu *Statystyka i sprawozdawczość* wybrać pozycję *Kursy*. Dostępne opcje eksportu danych zostały zaprezentowane na rysunku 1.7.

Rysunek 1.7. Bazy danych kursów walut na stronie NBP

Źródło: Opracowanie własne.

Jak widać na powyższym rysunku, dostępne są zestawienia archiwalne kursów walut w formatach csv oraz xls. Zarówno pierwszy, jak i drugi format pliku może być odczytany w narzędziu MS Excel. Rozważając przypadek pierwszy, należy jednak zastanowić się nad stroną kodową, w jakiej plik zostanie wczytany. Surowy plik w formacie csv charakteryzuje się brakiem ustalonej szerokości kolumn. Przyjęte zostały wartości domyślne, co spowodowało, że szerokość każdej z kolumn jest jednakowego rozmiaru. Zobrazowano to na rysunku 1.8 (plik *r1_kursy_walut_2015.csv*).

	A	B	C	D	E	F	G	H	I	J	K
1	data	1THB	1USD	1AUD	1HKD	1CAD	1NZD	1SGD	1EUR	100HUF	1C
2		bat (Tajlan dolar amer dolar austr dolar Hong dolar kana dolar nowc dolar singe euro forint (Węgrze									
3	20150102	0,1084	3,5725	2,9093	0,4606	3,0637	2,7681	2,6897	4,3078	1,3512	
4	20150105	0,1091	3,5975	2,9094	0,4638	3,0557	2,7601	2,6996	4,3008	1,3476	
5	20150107	0,1105	3,6375	2,9327	0,4691	3,0646	2,8183	2,7213	4,3115	1,349	
6	20150108	0,1109	3,6482	2,9595	0,4705	3,0879	2,8491	2,7285	4,2985	1,3535	
7	20150109	0,1102	3,6252	2,9438	0,4676	3,058	2,8327	2,7091	4,2837	1,3479	
8	20150112	0,1101	3,6218	2,9564	0,4671	3,0482	2,8134	2,7118	4,2782	1,3453	
9	20150113	0,1103	3,6252	2,9505	0,4675	3,0251	2,811	2,7122	4,2802	1,347	
10	20150114	0,1113	3,6525	2,971	0,4709	3,0494	2,8221	2,7342	4,2885	1,341	
11	20150115	0,1117	3,6588	2,9968	0,4718	3,0665	2,8414	2,7642	4,2897	1,3339	
12	20150116	0,1138	3,7174	3,053	0,4795	3,0987	2,8978	2,8066	4,322	1,3453	
13	20150119	0,114	3,7176	3,0504	0,4794	3,1063	2,8971	2,7917	4,3165	1,3565	
14	20150120	0,1141	3,7346	3,0638	0,4816	3,1201	2,8895	2,7923	4,3335	1,3601	
15	20150121	0,1146	3,7358	3,0565	0,4818	3,0879	2,8516	2,7966	4,3218	1,3686	
16	20150122	0,1135	3,6994	2,9991	0,4771	2,9986	2,7965	2,7753	4,2997	1,3628	
17	20150123	0,1156	3,7687	2,9997	0,4862	3,0368	2,8164	2,8063	4,2354	1,3601	

Rysunek 1.8. Dane średnich kursów walut za 2015 r. pobrane ze strony NBP

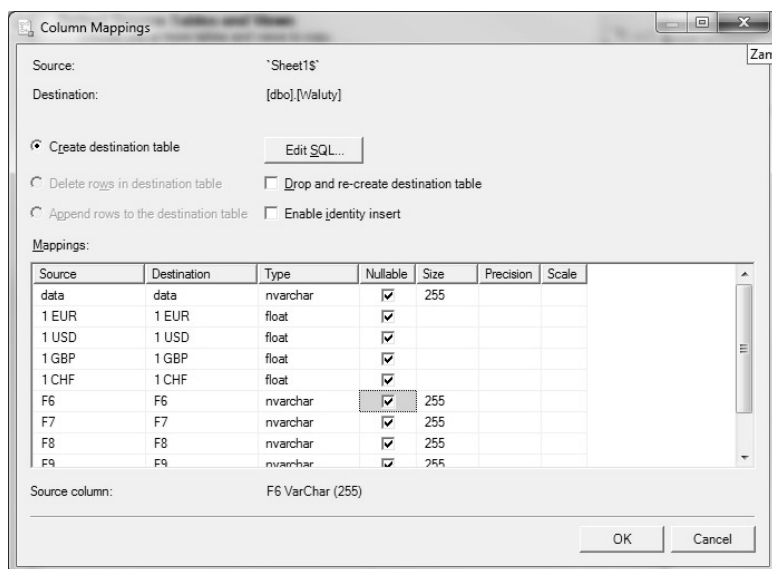
Źródło: Opracowanie własne.

Na powyższym rysunku można zauważyć, że liczba cyfr po przecinku nie zawsze jest taka sama. W przypadku waluty tajlandzkiej (symbol THB) 19 stycznia 2015 r. wartość wyniosła 0,114 zł za 1THB, podczas gdy w pozostałych dniach wartość ta zawsze podawana jest z czterema cyframi po przecinku. Dostarczając dane menedżerowi, powinno się dopilnować, aby liczba cyfr po przecinku była stała. Ponadto należy zadbać o prawidłowe wyświetlanie drugiego wiersza arkusza. Obecnie dane te nie mieszczą się w komórce.

Warto nadmienić, że po sformatowaniu danych pobranych w formacie csv powinno się je zachować w formacie xls lubxlsx. Takie formaty plików pozwalają na zapisywanie reguł formatowania komórek. Wówczas przy kolejnym otwarciu pliku nie będzie konieczne jego formatowanie od początku, co ma miejsce w przypadku pliku csv.

Edytowanie dużego zbioru danych w arkuszu kalkulacyjnym MS Excel w wielu przypadkach jest bardzo niewygodne. Wówczas stosuje się bazy danych. Przykładem bazy danych, która może stanowić bezpośrednie źródło danych dla arkusza MS Excel, jest Microsoft SQL Server. W celu wczytania danych z pliku MS Excel (*r1_kursy_walut.xlsx*) zawierającego waluty zastosowano narzędzie

MS SQL Server Import Export Wizard. Rozwiązanie to przygotowane dla Firmy X zostało zaprezentowane na rysunku 1.9.



Rysunek 1.9. Odzworowywanie nazw kolumn MS Excel w bazie danych MS SQL Server
Źródło: Opracowanie własne.

Na rysunku 1.9 widoczne są kolumny zawarte w arkuszu. W niektórych przypadkach podczas importowania danych może pojawić się sytuacja, w której poza standardowymi kolumnami (data i symbole walut) oznaczone zostały również puste kolumny symbolami: litera F z cyfrą oznaczającą numer kolumny.

Wczytanie zbioru danych skutkuje możliwością ich odczytu w narzędziu MS Excel. W tym celu na wstążce *Dane*, należy wybrać polecenie *Z innych źródeł* (grupa *Pobieranie danych zewnętrznych*), a następnie *Z programu SQL Server*. Po wpisaniu nazwy serwera i wybraniu bazy danych następuje pobranie danych. Zwykle wybierana jest opcja importowania danych do tabeli, co umożliwia później przeglądanie danych w określonych wymiarach. Podgląd taki znajduje się na rysunku 1.10.

	A	B	C	D	E	F
1	data	1 EUR	1 USD	1 GBP	1 CHF	
2	2015-06-03	4,1307	3,7108	5,6682	3,969	
3	2015-06-02	4,1374	3,7676	5,744	4,0019	
4	2015-06-01	4,126	3,785	5,7658	3,9984	
5	2015-05-29	4,1301	3,7671	5,759	3,991	
6	2015-05-28	4,1419	3,7858	5,8059	4,0014	
7	2015-05-27	4,1405	3,7906	5,8407	4,0014	
8	2015-05-26	4,1279	3,7898	5,8386	3,9864	
9	2015-05-25	4,1149	3,75	5,8063	3,9757	
10	2015-05-22	4,098	3,669	5,7447	3,9355	
11	2015-05-21	4.0836	3.6605	5.733	3.926	

Rysunek 1.10. Przeglądanie w MS Excel tabeli z bazy danych MS SQL Server

Źródło: Opracowanie własne.

Tak przygotowane dane są gotowe do dalszej obróbki. Należy zwrócić uwagę, że wczytywane dane są surowe i mogą zawierać błędy np. dotyczące niewłaściwego typu danych. Po wczytaniu takiego zbioru należy go zweryfikować, biorąc pod uwagę ewentualne usterki, co zostało szczegółowo opisane w rozdziale 2 niniejszego opracowania.

1.5. Przypadek 4 – pozyskiwanie danych giełdowych

Kolejnym rodzajem danych często wykorzystywanym w ramach analiz prowadzonych w firmach jest badanie indeksów giełdowych. Udostępniane przez Giełdę Papierów Wartościowych w Warszawie archiwa notowań pozwalają na włączenie ich do zintegrowanego systemu bazodanowego, pozwalającego ocenić warunki obecnie panujące na rynku. Dzięki danym archiwalnym można zbadać korelację pomiędzy sprzedażą produktów a nastrojami panującymi na giełdzie, czy też kursami walut, których sposób pozyskania opisano w poprzednim podrozdziale.

Firma X zaplanowała wdrożyć system, który będzie umożliwiał analizę indeksów na Giełdzie Papierów Wartościowych w Warszawie. Celem tego systemu miało być badanie, jak sytuacja na giełdzie wpływa na koniunkturę sprzedaży produktów. Miały być badane poszczególne firmy z branży zbliżonej do działalności Firmy X. Dane takie są ogólnodostępne na stronie internetowej gpw.pl. Archiwum notowań w postaci plików w formacie MS Excel można znaleźć na www.gpwinfostrafa.pl w sekcji *Archiwum Notowań*. Po wejściu na ten link przedsiębiorcy mają możliwość pobrania danych z następującymi pozycjami:

- kursy (otwarcia, minimalny, maksymalny, zamknięcia),
- liczba transakcji,

- obrót,
- wolumen obrotu,
- otwarte pozycje.

Podobnie jak w przypadku opisywanych wcześniej kursów walut również przetwarzanie danych giełdowych, ze względu na ich wielość, może się okazać trudne. Dlatego proponowane rozwiązanie zakłada wykorzystanie bazy danych MS Access w celu archiwizowania zbiorów danych. Baza ta pozwala na wczytywanie plików w formacie MS Excel i późniejsze ich pobieranie do arkusza kalkulacyjnego. Jest to bardzo istotne, gdyż wówczas w arkuszu znajdują się tylko te dane, których decydent potrzebuje.

W niniejszym przykładzie firma postanowiła analizować dane w podstawowym indeksie giełdowym, jakim jest WIG. Celem jest uzyskanie informacji o skłonności inwestorów do inwestowania na giełdzie w danym okresie. Założeniem jest doprowadzenie do sytuacji, w której możliwe będzie przeglądanie indeksu w następujących przekrojach:

- rok,
- miesiąc,
- dzień,
- symbol indeksu giełdowego,
- początek/koniec notowania.

Pierwszym etapem jest pobranie ze strony www.gpwinfostrefa.pl archiwum notowań. Należy w tym celu przejść do podstrony *Archiwum notowań* i wybrać opcję *wg instrumentu*. Na rysunku 1.11 zobrazowano przykład generowania i pobierania takiego pliku.

The image shows two screenshots of the 'ARCHIWUM NOTOWAŃ' (Archive of Quotes) interface on the website gpwinfostrefa.pl. The left screenshot shows 'KROK 1' (Step 1) where the user selects parameters for displaying the list of available instruments. It includes fields for 'Data od:' (2010-01-01) and 'Do:' (2015-07-30), a 'Generuj' button, and a dropdown menu for 'Rodzaj instrumentu:' (indeksy). The right screenshot shows 'KROK 2' (Step 2) where the user selects an instrument from the generated list. It includes a dropdown menu for 'Instrument:' (WIG) and a 'Generuj' button.

Rysunek 1.11. Dwuetapowe eksportowanie danych ze strony gpwinfostrefa.pl

Źródło: Opracowanie własne na podstawie [gpwinfostrefa.pl, dostęp dnia 30.07.2015].

W drugim etapie należało wybrać instrument giełdowy. Zgodnie z rysunkiem 1.11 w niniejszym studium przypadku był to indeks WIG. Tak wyeksportowany plik (*r1_indeks_WIG.xls*) lub grupę plików należało wczytać do bazy danych MS Access. Program ten umożliwił zaimportowanie wielu plików MS Excel, które w niniejszym studium przypadku zostaną zapisane jako niezależne tabele.

Na rysunku 1.12 pokazano wygląd tabeli wczytanej do MS Access.

Identyfikator	LP	Data	Typ Instrum.	Nazwa	ISIN	Kurs otwar.	Kurs maksyr.	Kurs minimum	Kurs zamkni.	Zmian
1	1	2014-01-02	1	WIG	PL9999999999	51670,42	51878,73	51494,42	51865,89	
2	2	2014-01-03	1	WIG	PL9999999999	51699,47	51785,32	51497,81	51497,81	
3	3	2014-01-07	1	WIG	PL9999999999	51213,21	51213,21	50412,37	50444,78	
4	4	2014-01-08	1	WIG	PL9999999999	50434,79	50610	50079,4	50482,93	
5	5	2014-01-09	1	WIG	PL9999999999	50409,08	50505,04	49753,03	49753,03	
6	6	2014-01-10	1	WIG	PL9999999999	49681,48	49796,5	49320,57	49796,5	
7	7	2014-01-13	1	WIG	PL9999999999	49943,58	50616,34	49927,54	50444,23	
8	8	2014-01-14	1	WIG	PL9999999999	50267,5	50766,49	50159,63	50605,70	

Rysunek 1.12. Integracja bazy danych MS Access z arkuszem kalkulacyjnym MS Excel

Źródło: Opracowanie własne.

Widoczna na rysunku 1.12 wstążka *Dane zewnętrzne* pozwala na wczytywanie do MS Access danych z różnego rodzaju źródeł. W opisywanym studium przypadku korzystano jedynie z danych zapisanych w formacie MS Excel. Tak przygotowane dane są możliwe do wczytania i integracji w ramach źródeł zewnętrznych w narzędziu MS Excel. W tym celu na wstążce *Dane* należy wybrać opcję *Z programu Access* (karta *Pobieranie danych zewnętrznych*), a następnie wskazać tabelę lub zbiór tabel zapisanych w bazie danych MS Access, które mają zostać zaimportowane.

Na rysunku 1.13. zaprezentowano dane wczytane do MS Excel, które zostały umieszczone w postaci tabeli przestawnej (plik *r1_dane_gieldowe.xlsx*).

Etykiety wierszy	Suma z Kurs maksymalny	Suma z Kurs minimalny
2015-07-23	51986,22	51438,32
2015-07-24	52125,02	51753,07
2015-07-27	51899,07	51287,52
2015-07-28	52132,64	51828,26
2015-07-29	52732,92	51943,42
Suma końcowa	260875,87	258250,59

Rysunek 1.13. Importowanie w arkuszu MS Excel danych nt. WIG pochodzących z bazy MS Access

Źródło: Opracowanie własne.

Więcej informacji na temat tabel przestawnych znajduje się w rozdziale 5 niniejszego opracowania. Mechanizmy pozwalające na integrację danych pochodzących z wielu źródeł przedstawiono w rozdziale 6.

1.6. Przypadek 5 – pozyskiwanie danych z bazy Eurostat na temat wysokości wynagrodzeń i poziomu wykształcenia

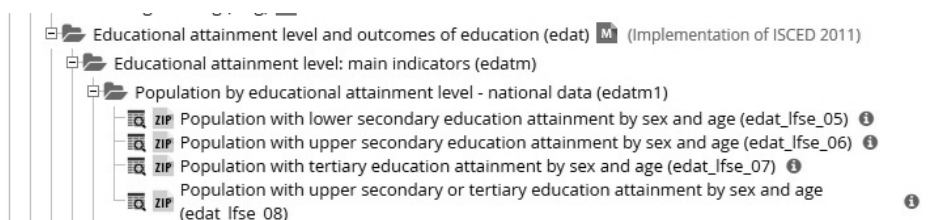
Ważkim tematem dotyczącym funkcjonowania firm jest analiza wynagrodzeń osób. Firma usługowa, która prowadzi działalność w skali międzynarodowej, stara się zwykle lokalizować swoje centra w krajach, w których koszty pracy są niskie. Nie jest to jedyne kryterium wyboru lokalizacji. Równie istotny jest dostęp do wysoko wykwalifikowanej kadry.

W niniejszym podrozdziale przedstawiono przykład Firmy X, która działa w Europie. Celem planowanego podsystemu jest stałe monitorowanie rynku pracy oraz wykształcenia osób w różnych krajach, aby rekrutować ich do pracy w danym regionie. Firma X do analiz wykorzystuje ogólnodostępne dane pochodzące z bazy danych Eurostatu, którego celem jest dostarczanie wysokiej jakości danych statystycznych. Dane te zostały pobrane w postaci wskaźników dla poszczególnych krajów, odpowiednio dotyczących poziomu wykształcenia i wynagrodzeń. Dane te znajdują się na stronie: <http://ec.europa.eu/eurostat/data/database>.

Z bazy danych wybrano następujące zbiory danych:

- Osoby z wyższym wykształceniem (*Population with tertiary education attainment by sex and age* w sekcji *Education and Training*);
- Roczne wynagrodzenie brutto (*Average annual gross earnings by occupation* w sekcji *Labour Market, Gross Earnings: historical data*).

Na rysunku 1.14 przedstawiono wybrane pozycje bazy danych Eurostat.



Rysunek 1.14. Baza danych Eurostat – pozyskiwanie danych

Źródło: Opracowanie własne.

Celem zapisania danych do dalszej obróbki należy określić format eksportowanych danych, po wcześniejszym wybraniu odpowiedniej tablicy w bazie da-

nych Eurostat. Istnieje również możliwość zapisania danych w postaci plikowej bez wyświetlania tablicy (por. rysunek 1.14).

Istotnym wyzwaniem, jakie zostało postawione przed analitykami danych, było zintegrowanie zbiorów danych. W tym celu pobrano pliki w formacie MS Excel (pliki *r1_wyzsze_wyksztalzenie.xlsx* oraz *r1_wynagrodzenia_brutto.xlsx*) i wykorzystano dwa arkusze, aby je przechować w celu wykonania późniejszych analiz. Na rysunku 1.15 zaprezentowano zawartość pliku z wynagrodzeniami brutto.

GEO/TIME	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Belgium	33 873,0
Bulgaria	1 256,8	1 411,2	1 517,3	1 638,1	1 737,8	1 859,9	2 056,4	2 269,1	2 699,7	3 415,0
Denmark	39 532,9	40 996,5	41 694,8	43 577,3	44 692,0	46 122,0	47 529,3	48 307,3	50 578,1	52 703,1
Germany (until 1990 former territo	34 200,0	34 700,0	35 400,0	36 500,0	37 300,0	38 000,0	38 600,0	39 066,0	40 100,0	41 200,0
Ireland	40 944,0	.	.	.	40 932,0	.
Greece	14 882,5	15 788,4	16 630,0	17 509,8	18 122,2
Spain	.	.	17 873,8	18 580,3	19 367,7	20 018,6	20 593,9	21 724,7	22 176,5	.
Cyprus	16 574,2	17 687,1	18 361,4	19 396,5	20 032,5	20 541,2	21 842,1	22 836,4	23 662,9	25 579,0
Hungary	6 929,3	7 395,6	8 255,5	8 280,0	9 338,8	.
Malta	.	13 505,8	14 267,0	14 442,3	14 294,6	14 522,0	15 011,0	15 593,0	16 027,0	16 313,0
Austria	.	.	33 013,9	34 262,0	35 186,0	35 315,0	36 533,0	37 372,0	38 458,7	39 761,0
Poland	6 278,7
Portugal	.	12 572,6	13 309,0	13 336,4	13 491,4	13 786,5	14 068,4	16 221,6	17 201,6	17 753
Romania	2 429,0	3 218,0	3 910,6	5 044,2	5 760,3
Slovakia	3 005,0	3 426,7	3 619,9	4 263,9	4 644,2	5 388,5	6 047,2	6 706,9	8 031,3	.
Finland	.	26 584,8	27 897,2	29 060,0	30 054,0	31 055,0	32 377,0	33 069,0	34 738,0	36 625,0
Sweden	.	31 500,9	30 271,1	30 957,5	32 034,0	32 344,2	32 902,3	33 915,3	35 534,0	35 442,1
United Kingdom	.	.	38 702,0	.	37 309,8	39 768,1	.	42 654,7	43 675,4	.
Iceland	32 312,0	37 640,9	34 099,7	35 233,0	36 914,7	41 362,0	56 768,5	55 891,1	64 042,0	39 977,8

Rysunek 1.15. Baza danych Eurostat – dane na temat wynagrodzeń w arkuszu MS Excel

Źródło: Opracowanie własne.

Należy zwrócić uwagę, że zaprezentowany na rysunku 1.15 zbiór danych posiada wartości brakujące, oznaczone dwukropkiem. Jeżeli wykonywane są działania arytmetyczne w MS Excel, na przykład sumowania poprzez dodawanie kolejnych komórek, takiego rodzaju dane mogą spowodować wystąpienie błędu. Na rysunku 1.16 zaprezentowano taką sytuację.

	GEO/TIME	1999	2000	2001	Formuła = Bnn + Cnn + Dnn	Formuła = SUMA (Bnn; Cnn; Dnn)
12	Belgium	.	.	.	#ARG!	0
13	Bulgaria	1 256,8	1 411,2	1 517,3	4185,3	4185,3
14	Denmark	39 532,9	40 996,5	41 694,8	122224,2	122224,2
15	Germany (until 1990 former territo	34 200,0	34 700,0	35 400,0	104300	104300
16	Ireland	.	.	.	#ARG!	0
17	Greece	14 882,5	15 788,4	16 630,0	47300,9	47300,9
18	Spain	.	.	17 873,8	#ARG!	17873,8
19	Cyprus	16 574,2	17 687,1	18 361,4	52622,7	52622,7
20	Hungary	.	.	.	#ARG!	0
21	Malta	.	13 505,8	14 267,0	#ARG!	27772,8
22	Austria	.	.	33 013,9	#ARG!	33013,9

Rysunek 1.16. Błędy w przetwarzanych danych na temat wynagrodzeń

Źródło: Opracowanie własne.

Przykładowo w wierszu 13 opisującym wynagrodzenia w Belgii pierwsza formuła ma następującą postać: =B13+C13+D13. Generuje ona błędy, gdyż

zastosowano dodawanie wartości tekstowej, jaką jest dwukropek powiązany operatorem dodawania (symbol +) z wartościami liczbowymi. Aby uniknąć takiej sytuacji, należy zastosować funkcję, która nie jest wrażliwa na występowanie wartości tekstowych. Funkcją taką jest SUMA lub SUM w angielskiej wersji programu MS Excel. Jak widać na rysunku 1.16, funkcja ta nie powoduje wygenerowania błędu w przypadku sumowania wartości tekstowych. Dla wiersza 13 ma ona postać: =SUMA(B13;C13;D13).

Wykorzystywanie tego rodzaju funkcji niesie jednak zagrożenie, że wartości liczbowe zapisane jako tekstowe nie będą wzięte pod uwagę. Są to wartości poprzedzone znakiem apostrofu, o czym była mowa w niniejszym rozdziale.

W opisywanym studium przypadku firma dodatkowo wykorzystuje informacje na temat wykształcenia osób w krajach europejskich, aby móc inwestować w tych regionach. W tym celu pobrano arkusz z danymi na temat ludności z wyższym wykształceniem (plik *r1_wyzsze_wyksztalzenie.xlsx*), zaprezentowany na rysunku 1.17.

	GEO/TIME	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
12	European U	19,6	20,0	20,5	21,2	22,0	22,8	23,7	24,6	25,4	26,0
14	European U	19,7	20,1	20,6	21,3	22,1	22,8	23,8	24,7	25,5	26,1
15	European U	21,4	21,8	22,3	22,9	23,7	24,3	25,2	26,1	26,9	27,3
16	Euro area (1	20,1	20,3	20,8	21,5	22,2	22,6	23,5	24,2	25,0	25,3
17	Euro area (1	20,1	20,3	20,8	21,5	22,1	22,6	23,4	24,2	24,9	25,3
18	Euro area (1	20,1	20,3	20,8	21,5	22,1	22,6	23,4	24,2	24,9	25,3
19	Belgium	27,2	27,9	28,1	28,4	29,4	30,7	30,4	31,3	31,5	32,6
20	Bulgaria	17,8	18,2	18,5	18,9	19,2	19,4	20,1	20,7	22,2	23,6
21	Czech Rept	11,0	11,4	11,6	12,4	13,4	14,5	15,8	17,0	18,1	19,1
22	Denmark	28,5	29,3	26,0	26,3	26,9	27,5	27,9	28,6	29,1	29,4

Rysunek 1.17. Baza danych Eurostat – dane na temat wykształcenia ludności

Źródło: Opracowanie własne.

Należy zwrócić uwagę, że wartości nieaddytywne, czyli takie, które nie powinny się sumować, są zapisywane jako tekst. Widać to na przykładzie lat na rysunku 1.17. Komórki z wartością roku nie posiadają domyślnego formatowania (w lewym górnym rogu znajduje się symbol zielonego trójkąta). Prawidłowym formatem dla tej wartości byłby typ liczbowy. Jak wspomniano, nie ma to jednak większego uzasadnienia, gdyż wartości roku należą nie do danych, lecz do tzw. metadanych, czyli wartości opisujących, co znajduje się na przecięciu kolumn i wierszy.

Podsumowując, dostępność licznych baz danych makroekonomicznych na stronie internetowej Eurostatu powoduje, że decydenci mają możliwość poszerzenia swojego zakresu wiedzy dzięki dostępności danych porównywalnych w skali międzynarodowej.

1.7. Możliwe zastosowania MS Excel w biznesie w formie zadań

W celu zobrazowania możliwości pozyskiwania danych poprzez narzędzie MS Excel można posłużyć się poniższymi zadaniami:

1. Wykorzystując stronę Eurostat, Firma X może pozyskać dane dotyczące kosztów i przychodów funkcjonowania firm w Unii Europejskiej w ostatnich 10 latach. Należy zatem pobrać te dane ze strony internetowej, upewniając się, że zawarte zostały koszty i przychody działalności przedsiębiorstw.

2. Firma X planuje zlokalizować swoje oddziały w pięciu województwach w Polsce. Jako kryterium wyboru przyjęła odsetek osób z wykształceniem wyższym w ostatnim możliwym roku referencyjnym. Rozwiązaniem dla Firmy X jest pozyskanie oficjalnych danych statystycznych. Należy zatem odnaleźć właściwe dane i pobrać je do MS Excel, wykorzystując przy tym relacyjną bazę danych MS SQL Server.

3. Firma X chce rozszerzyć swoją działalność eksportową, oferując produkty do przechowywania warzyw i owoców do 5 krajów, do których tego rodzaju towary są eksportowane w największej skali. Należy przygotować zbiór danych w MS Excel dla 5 ostatnich lat referencyjnych.

4. Firma X (w ramach przypadku 3) ma zamiar dodać kolejne dane do arkusza walut. Znaleźć należy możliwe źródła danych dla notowań walut i zasilić nimi arkusz danych.

5. Firma X ma zamiar pozyskać informacje na temat możliwości wymiany handlowej z Wielką Brytanią. Wykorzystując bazę danych Eurostatu, należy odnaleźć wszelkiego rodzaju dane, które informują na temat wielkości eksportu i importu pomiędzy Wielką Brytanią a Polską.

Zakończenie

Zaprezentowane w niniejszym rozdziale studia przypadków jednoznacznie wskazują, że MS Excel może być narzędziem powszechnie wykorzystywanym w przedsiębiorstwie do pozyskiwania i wstępnego przetwarzania danych. Dzięki licznym narzędziom dostępnym w tym programie firma jest w stanie analizować dane pochodzące z wielu źródeł. Przykładem są zaprezentowane dane dotyczące notowań giełdowych (nastroje inwestorów), walut (kondycja gospodarki) i sprzedaży. Jest to podstawowy zestaw danych, który może zostać wykorzystany do analiz statystycznych, takich jak korelacja pomiędzy różnego rodzaju wartościami. Firma eksportująca towary może sprawdzić przykładowo, czy był związek pomiędzy kursem waluty a liczbą eksportowanych do danego kraju towarów.

Firmy, które są zainteresowane pozyskiwaniem wiarygodnych danych, mają do wyboru wiele różnego rodzaju zewnętrznych baz danych. Ze względu na różny format tych danych istotne jest zapewnienie dostępności wielu interfejsów dostępowych. Spektrum takich narzędzi znajduje się między innymi w arkuszu kalkulacyjnym MS Excel, który został wykorzystany w opisywanych studiach przypadków. Aby uniknąć błędów w wynikach podczas późniejszego przetwarzania danych, trzeba jednak odpowiednio przygotować te dane, co nie było jednak przedmiotem rozważań w niniejszym rozdziale. Informacje, w jaki sposób dane powinny być przetwarzane, znajduje się w kolejnym rozdziale.

Obecnie firmy coraz częściej będą sięgać po tego rodzaju dane, ze względu na bardziej rozbudowane możliwości analityczne popularnych narzędzi biurowych. Szeroka dostępność narzędzi analitycznych oraz ich znajomość przez szeregowych pracowników może przyczynić się do podniesienia konkurencyjności przedsiębiorstw.

Bibliografia

1. Duda J. (2012), *Business Intelligence and NOSQL Databases*, „Information Systems in Management”, Vol. 1, No. 1.
2. Kim Y.I., Ji Y.K., Park S. (2014), *Big Text Data Clustering using Class Labels and Semantic Feature Based on Hadoop of Cloud Computing*, „International Journal of Software Engineering & Its Applications”, Vol. 8, Issue 4.
3. Kuiler E. W. (2014), *From Big Data to Knowledge: An Ontological Approach to Big Data Analytics*, w: „Review of Policy Research”, Vol. 31, Issue 4.
4. Kwon H., Lee N., Shin B. (2014), *Data quality management, data usage experience and acquisition intention of big data analytics*, „International Journal of Information Management”, Vol. 34, Issue 3.
5. Mousavizadeh M., Harden G., Ryan S., Windsor J. (2015), *Knowledge Management and the Creation of Business Value*, „Journal of Computer Information Systems”, Vol. 55, No. 4.
6. Tripathy A., Das K. (2011), *A Descriptive Approach towards Data Warehouse and OLAP Technology: An Overview*, w: *Computer Networks & Information Technologies, Computer Networks and Information Technologies*.
7. Wood C. (1999), *OLEDDB and ODBC developer's guide*, M&T Books.
8. Wrycza S. (2010), *Informatyka Ekonomiczna. Podręcznik akademicki*, PWE, Warszawa.
9. <http://ec.europa.eu/eurostat/data/database>, dostęp dnia 1.06.2015.
10. <http://stat.gov.pl>, dostęp dnia 28.05.2015.
11. <http://www.gpw.pl>, dostęp dnia 30.07.2015.
12. <http://www.gpwinfostrefa.pl>, dostęp dnia 30.07.2015.
13. <http://www.nbp.pl>, dostęp dnia 3.06.2015.
14. [https://msdn.microsoft.com/en-us/library/ms712640\(v=vs.85\).aspx](https://msdn.microsoft.com/en-us/library/ms712640(v=vs.85).aspx), dostęp dnia 19.07.2015.

Rozdział 2

Przygotowanie danych do analizy

Dariusz Kralewski, Monika Woźniak

Wstęp

Dane pozyskane ze źródeł zewnętrznych należy odpowiednio przygotować do analizy, dbając o ich wysoką jakość. Według Platona, zgodnie z kategorią filozoficzną, jakość to „pewien stopień doskonałości”. Inaczej jakość można określić jako zgodność z celem, zgodność ze specyfikacją (zero braków), stopień doskonałości wyrobu lub usługi, zespół cech i charakterystyk wyrobu lub usługi, które noszą w sobie zdolność zaspokojenia określonej potrzeby [EIPA, 2013].

Zatem jakość danych należy postrzegać jako ogół cech ukierunkowany na zdolność do zaspokojenia wyrażonych i ukrytych potrzeb [Harper, 2015]. W modelu jakości danych jakość definiowana jest jako ogólny wskaźnik stosowany do ustalenia przydatności zbioru danych do wykorzystania w określonym celu. Dane są wysokiej jakości, jeżeli nadają się do użycia zgodnie z przeznaczeniem w zakresie działania, podejmowania decyzji i planowania. Dane nadają się do zastosowania zgodnie z przeznaczeniem, jeżeli nie zawierają błędów i posiadają pożądane cechy [Redman, 2001].

Norma ISO dokładnie opisuje komponenty jakości danych. Są to: kompletność, spójność logiczna, dokładność, dokładność czasowa oraz dokładność tematyczna [ISO/TS 8000-1:2011]. Kompletność to obecność obiektów, ich atrybutów i związków (nadmiar, niedomiar danych). Spójność logiczna (pojęciowa, dziedziny, topologiczna, formatu) określa stopień zgodności z logicznymi regułami struktury danych, atrybutów i związków. Dokładność odnosi się do precyzji konstrukcji obiektów danych. Dokładność czasowa definiowana jest jako dokładność czasowych (związanych z czasem) atrybutów i związków obiektów. Dokładność tematyczna dotyczy dokładności atrybutów ilościowych, poprawności atrybutów nieilościowych oraz klasyfikacji obiektów i ich związków.

Dane są podatne na różnego rodzaju błędy. Mogą być one spowodowane czynnikiem ludzkim poprzez działania użytkowników, wady aplikacji lub wielokrotne migrowanie i integrowanie danych. Mogą powstać wskutek czynników zewnętrznych, takich jak awarie sprzętowe lub samoistne starzenie się danych [McGilvray, 2008].

Zatem jakość danych jest pojęciem względnym. Nie można zastosować jednego kryterium oceny do wszystkich możliwych sytuacji. W celu zbadania jakości

danych potrzebna jest analiza i poznanie charakteru danych oraz zapoznanie się z przeznaczeniem danych i procesów, jakim są one poddawane.

Aby zapewnić wysoką jakość danych, należy je odpowiednio uporządkować i sformatować, tak aby stały się czytelne dla odbiorcy [Alexander, Decker, 2014]. W niniejszym rozdziale przedstawione zostały m.in. mechanizmy formatowania danych. Sformatowane i uporządkowane dane będą punktem wyjściowym do analizy podejmowanej w kolejnych rozdziałach.

Celem bieżącego rozdziału jest przegląd oraz prezentacja możliwości przygotowania danych biznesowych do późniejszej analizy, a tym samym podniesienia ich jakości. Wszystkie opisane mechanizmy przygotowania danych do analizy mają zastosowanie w praktyce. Dla lepszego zilustrowania ich użyteczności, poniżej przedstawiono studia przypadków.

2.1. Przypadek 1 – przygotowanie danych na temat eksportu i importu z wykorzystaniem mechanizmów formatowania

Niniejsze studium przypadku odnosi się do Firmy X prowadzącej handel międzynarodowy. W poprzednim rozdziale, w ramach pierwszego studium przypadku, pozyskano dane z bazy zawierającej dane statystyczne na temat importu i eksportu towarów i usług (plik *r1_import_i_eksport_no.xlsx*). Zbiór ten znajduje się w czterech osobnych arkuszach, odpowiednio obrazujących wielkość eksportu oraz importu w latach 2013–2014. Dane takie przed dalszą obróbką należy ujednoczyć i odpowiednio sformatować. Zaimportowany zbiór danych zawiera liczne błędy, które powinny zostać usunięte. Zatem w fazie przygotowania danych do analizy należy:

- usunąć scalanie komórek;
- wyeliminować przesunięcia kolumn (niektóre dane są rozbite w wielu kolumnach);
- usunąć wiersze podsumowujące w wybranych arkuszach, co spowoduje, że takie dane będzie można agregować;
- sformatować odpowiednio komórki – nadać format walutowy kolumnom: *Wartość w zł*, *Wartość w euro*, *Wartość w dol.*;
- sformatować całość danych jako tabelę, tak aby otrzymać pożądane obramowanie.

Wszystkie operacje powinno się wykonać w arkuszach *EKSPORT_2013*, *EKSPORT_2014*, *IMPORT_2013* i *IMPORT_2014*. Rezultat, na przykładzie pierwszego z nich, pokazany został na rysunku 2.1.

	A	B	C	D	E	F	G	H	I
1	EKSPORT - rok 2013, 4 kwartały								
2									
3	Kod	Nazwa towaru	Kraj	Symb. kraju	Jedn miary	Ilość	Wartość w zł	Wartość w euro	Wartość w dol.
4	0808	Jabłka, gruszek i pigwy, świeże	Białoruś	BY		208 698 506,00 zł	€ 50 203 176,00	\$ 66 576 470,00	
5	0808	Jabłka, gruszek i pigwy, świeże	Moldawia	MD		189 401,00 zł	€ 46 422,00	\$ 59 421,00	
6	0808	Jabłka, gruszek i pigwy, świeże	Rosja	RU		1 134 067 867,00 zł	€ 273 036 642,00	\$ 361 360 783,00	
7	0808	Jabłka, gruszek i pigwy, świeże	Ukraina	UA		83 863 486,00 zł	€ 20 210 154,00	\$ 26 591 959,00	
8	0808	Jabłka, gruszek i pigwy, świeże	Algieria	DZ		7 547 080,00 zł	€ 1 813 110,00	\$ 2 402 168,00	
9	0808	Jabłka, gruszek i pigwy, świeże	Antarktyda	AQ		1 125,00 zł	€ 265,00	\$ 355,00	
10	0808	Jabłka, gruszek i pigwy, świeże	Arabia Saudyjska	SA		98 517,00 zł	€ 23 693,00	\$ 31 214,00	
11	0808	Jabłka, gruszek i pigwy, świeże	Azerbejdżan	AZ		527 857,00 zł	€ 127 011,00	\$ 169 424,00	
12	0808	Jabłka, gruszek i pigwy, świeże	Bahrajn	BH		72 479,00 zł	€ 17 427,00	\$ 23 234,00	
13	0808	Jabłka, gruszek i pigwy, świeże	Bośnia i Hercegowina	BA		4 889 882,00 zł	€ 1 182 053,00	\$ 1 561 255,00	
14	0808	Jabłka, gruszek i pigwy, świeże	Czarnogóra	ME		158 711,00 zł	€ 38 040,00	\$ 49 494,00	
15	0808	Jabłka, gruszek i pigwy, świeże	Egipt	EG		3 094 810,00 zł	€ 746 336,00	\$ 998 175,00	
16	0808	Jabłka, gruszek i pigwy, świeże	Gruzja	GE		766 456,00 zł	€ 184 663,00	\$ 243 821,00	
17	0808	Jabłka, gruszek i pigwy, świeże	Indie	IN		99 945,00 zł	€ 24 205,00	\$ 32 429,00	
18	0808	Jabłka, gruszek i pigwy, świeże	Irak	IQ		26 295,00 zł	€ 6 284,00	\$ 8 497,00	
19	0808	Jabłka, gruszek i pigwy, świeże	Jordania	JO		997 124,00 zł	€ 240 284,00	\$ 320 198,00	
20	0808	Jabłka, gruszek i pigwy, świeże	Kazachstan	KZ		68 030 787,00 zł	€ 16 374 171,00	\$ 21 698 312,00	
21	0808	Jabłka, gruszek i pigwy, świeże	Kosowo	XK		203 900,00 zł	€ 48 960,00	\$ 63 628,00	
22	0808	Jabłka, gruszek i pigwy, świeże	Libia	LY		1 845 125,00 zł	€ 444 676,00	\$ 591 037,00	
23	0808	Jabłka, gruszek i pigwy, świeże	Majotta	YT		32 476,00 zł	€ 7 806,00	\$ 10 475,00	
24	0808	Jabłka, gruszek i pigwy, świeże	Maroko	MA		109 292,00 zł	€ 26 654,00	\$ 35 425,00	
25	0808	Jabłka, gruszek i pigwy, świeże	Mongolia	MN		991 741,00 zł	€ 237 985,00	\$ 319 242,00	
26	0808	Jabłka, gruszek i pigwy, świeże	Nigeria	NG		51 554,00 zł	€ 12 330,00	\$ 15 950,00	
27	0808	Jabłka, gruszek i pigwy, świeże	QQ w ramach handlu z kr. trzecimi	QS		5 025,00 zł	€ 1 187,00	\$ 1 583,00	

Rysunek 2.1. Rezultat przygotowania danych dotyczących eksportu w 2013 roku do analizy
Źródło: Opracowanie własne.

W komórkach arkusza kalkulacyjnego znajdują się dane, które należy przetworzyć. Wyróżnia się trzy rodzaje danych. Są to:

- dane liczbowe,
- dane tekstowe,
- wzory i funkcje nazywane formułami.

Każdy prezentowany arkusz powinien być estetyczny i czytelny. Wygląd arkusza powinien skłaniać do jego analizy, a funkcjonalna forma nakierowywać analizującego na ważne informacje [Nussbaumer Knaflic, 2015]. Narzędzia służące do formatowania znajdują się na wstążce *Narzędzia główne*.

Do najczęściej wykonywanych czynności związanych z formatowaniem należą:

- obramowanie komórek,
- wypełnienie komórek kolorem,
- ustalenie koloru czcionki, jej wielkości i kształtu, pogrubienie itd.,
- wyrównanie zawartości komórki do strony lewej, prawej lub wyśrodkowanie, i to zarówno w poziomie, jak i w pionie,
- określenie, czy dłuższy tekst w komórce ma być zawijany,
- określenie sposobu wyświetlania liczb w komórce itd.

Arkusz kalkulacyjny MS Excel oferuje szeroką gamę wbudowanych formatów standardowych, takich jak: liczbowe, walutowe, księgowo, daty i czasu itp. Każdy z tych formatów można dodatkowo przystosowywać do potrzeb firmy (liczba wyświetlanych miejsc po przecinku w formacie liczbowym, znak waluty

dla formatu księgowego, sposób prezentacji liczb ujemnych itd.). Twórcy MS Excel dodali też kilka formatów specjalnych (kod pocztowy, numer telefonu, numer PESEL i NIP).

Zdarza się jednak, że w firmie występuje potrzeba wykorzystania formatu wykraczającego poza standardowe ramy. Użyć można wtedy tzw. formatowania niestandardowego czyli narzędzia umożliwiającego zdefiniowanie wyglądu danych w komórce. Struktura definicji formatu niestandardowego składa się z czterech sekcji rozdzielonych średnikami według następującej kolejności: *format_liczby_dodatniej;format_liczby_ujemnej;format_zera;format_tekstu*.

Znaki wykorzystywane podczas budowy formatów niestandardowych są dokładnie opisane w pomocy programu MS Excel. W tabeli 2.1 zaprezentowano przykłady definiowania formatów niestandardowych, które stosowane są podczas formatowania firmowej tabeli.

Tabela 2.1. Przykłady formatów niestandardowych

Format	Liczba dodatnia (1000)	Liczba ujemna (-1000)	Zero (0)	Tekst (Tekst)
##0,00	1 000,00	-1 000,00	0,00	Tekst
##0;[Czerwony]-##0	1 000	-1 000	0	Tekst
zł ##0;[Czerwony]zł -##0;zł 0,000	zł 1 000	zł -1 000	zł 0,000	Tekst
zł ##0,0;[Czerwony]zł -##0,0;zł 0,000; „przykładowy” @	zł 1 000,0	zł -1 000,0	zł 0,000	przykładowy tekst

Czcionkę w kolorze czerwonym zastąpiono w niniejszej publikacji czcionką pogrubioną.

Źródło: Opracowanie własne

Przykłady zaprezentowane w tabeli obrazują wykorzystanie znaków budujących format niestandardowy.

Wyniki przygotowania danych dotyczących importu i eksportu zawiera plik *r2_import_i_eksport_o.xlsx*.

2.2. Przypadek 2 – przygotowanie danych do analizy z wykorzystaniem formuł tekstowych

Wygenerowany w Firmie X (rozdział 1, przypadek 2) plik prezentujący dane dotyczące sprzedaży (plik *r1_dane_sprzedaży_no.csv*) zawiera dużą liczbę błędów, które wymagają usunięcia przed dalszą obróbką danych. Zagadnienia, jakie analityk danych musi rozpatrzyć w takim zbiorze, obejmują:

- strony kodowe mogą być niezgodne z systemem – należy wówczas na etapie otwierania zbioru danych wybrać właściwą stronę kodową;

- należy zwracać uwagę na sposób zapisu liczb – niekiedy kodowanie jest niejednolite – np. liczby po kropce, po przecinku itd.;
- obok miast znajdują się skróty nazw województwa – należy je pozyskać;
- z kolumn należy wybrać jak najwięcej atrybutów, np. rok, miesiąc, dzień itd.

Podsumowując, analityk przed rozpoczęciem analizy zbioru danych jest zobowiązany do jego przygotowania, aby spełniał stawiane w firmie wymagania jakościowe.

Celem niniejszego studium przypadku jest ukazanie etapów przygotowania zbioru danych do analiz na przykładzie danych dotyczących sprzedaży produktów w Firmie X. Otwarcie pliku csv bezpośrednio w MS Excel spowodowało błędne kodowanie polskich znaków (rysunek 2.2).

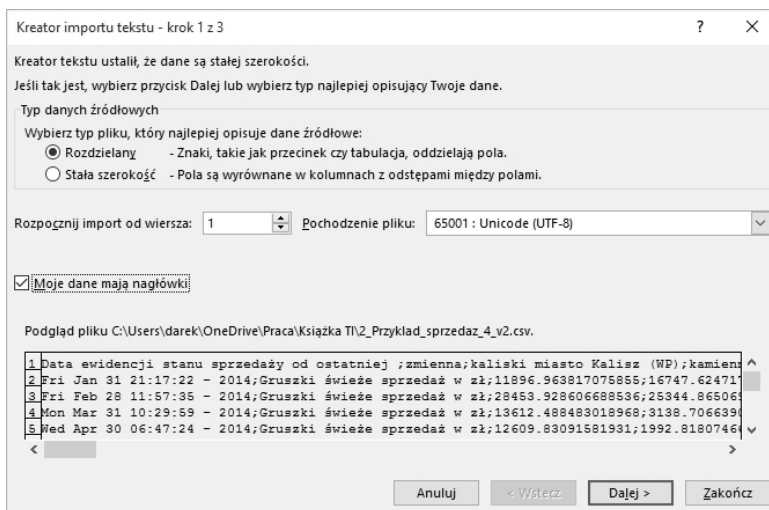
	A	B	C
1	Data ewidencji stanu sprzedaży od ostatniej	zmienna	kaliski miasto Kalisz (WP)
2	Fri Jan 31 21:17:22 - 2014	Gruszki świeże sprzedał w zł,	11896.963817075855
3	Fri Feb 28 11:57:35 - 2014	Gruszki świeże sprzedał w zł,	28453.928606688536
4	Mon Mar 31 10:29:59 - 2014	Gruszki świeże sprzedał w zł,	13612.488483018968
5	Wed Apr 30 06:47:24 - 2014	Gruszki świeże sprzedał w zł,	12609.83091581931
6	Sat May 31 00:25:48 - 2014	Gruszki świeże sprzedał w zł,	12031.149889363096
7	Mon Jun 30 21:57:22 - 2014	Gruszki świeże sprzedał w zł,	9224.673597996802
8	Thu Jul 31 14:00:21 - 2014	Gruszki świeże sprzedał w zł,	12874.537711916211
9	Sun Aug 31 14:17:44 - 2014	Gruszki świeże sprzedał w zł,	20535.896862520134

Rysunek 2.2. Niepoprawny format polskich znaków, rezultat otwarcia pliku csv bezpośrednio w MS Excel

Źródło: Opracowanie własne.

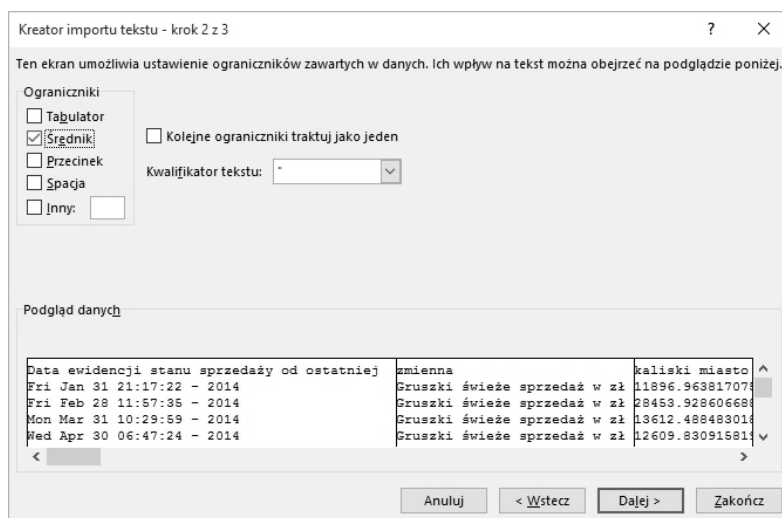
Aby wyeliminować ten błąd, należało za pośrednictwem polecenia *Pobieranie danych zewnętrznych, Z tekstu* (wstążka *Dane*) zaimportować plik csv. Istotne jest określenie prawidłowych parametrów importowania (rysunek 2.3), a przede wszystkim wskazanie sposobu rozdzielania tekstu na kolumny (w tym przypadku tekst jest rozdzielany za pomocą znaków specjalnych), sposobu kodowania (w opisywanym przypadku to Unicode – UTF-8) oraz zdefiniowanie, że dane mają nagłówki.

W przypadku gdy rozdzielnikiem tekstu jest znak specjalny, istotne jest poprawne wskazanie rodzaju separatora (w tym przypadku jest to znak średnika). Podgląd umożliwia zweryfikowanie poprawności (rysunek 2.4). Problemy mogą się pojawić, gdy znak separatora pojawia się dodatkowo w roli zwykłego znaku. Przykładowo, separatorem może być przecinek, a w danych występują liczby dziesiętne zapisane po przecinku (zgodnie z polską notacją), a nie po kropce (zgodnie z amerykańską notacją).



Rysunek 2.3. Okno dialogowe importu danych z pliku tekstowego

Źródło: Opracowanie własne.



Rysunek 2.4. Okno dialogowe importu tekstu – znak rozdzielnika i podgląd

Źródło: Opracowanie własne.

Wszystkie dane zaimportowane zostały w formacie tekstowym, zatem przed dalszą analizą należało zdefiniować poprawne formatowanie.

Bardzo często, zwłaszcza gdy firma importuje dane z zewnętrznych systemów, zdarza się, że format liczb zgodny jest z amerykańską notacją, czyli jako separator tysięcy używany jest przecinek, a jako symbol dziesiętny – kropka (np.: 1,234.56 zamiast 1234,56). Podczas importu danych do MS Excel format zostaje

źródłowy. Liczby zapisane zgodnie z amerykańską notacją w MS Excel traktowane są jako tekst. W przypadku danych dotyczących sprzedaży w Firmie X wszystkie wartości są niepoprawnie sformatowane. Należało zatem (wykorzystując narzędzie *Znajdowanie i zamienianie*) zamienić kropki na przecinki we wszystkich komórkach, które powinny mieć format liczbowy.

Formuły w arkuszu kalkulacyjnym służą do wykonywania operacji na liczbach i tekście. Wszystkie funkcje dostępne w arkuszu kalkulacyjnym uporządkowane są według kategorii (więcej na ten temat w rozdziale 3). Na etapie oczyszczania danych i przygotowania do analizy często wykorzystuje się funkcje tekstowe [Walkenbach, 2013]. Dokładny opis funkcji tekstowych znajduje się w dokumentacji. Poniżej zaprezentowane zostały te, które najczęściej stosowane są w przygotowywaniu danych do analiz biznesowych.

Funkcja ZŁĄCZ.TEKSTY (ang. CONCATENATE) łączy kilka ciągów tekstowych w jeden. Za jej pomocą można połączyć teksty znajdujące się w różnych kolumnach. Elementy tekstowe można także łączyć za pomocą operatora obliczeń w postaci „i” (&). Na przykład funkcja =A1&B1 zwraca taką samą wartość jak =ZŁĄCZ.TEKSTY(A1;B1).

Funkcja PORÓWNAJ (ang. EXACT) porównuje dwa teksty i zwraca wartość PRAWDA, jeśli są dokładnie takie same; w przeciwnym przypadku zwraca wartość FAŁSZ. Funkcja PORÓWNAJ uwzględnia wielkość liter, ale ignoruje różnice w formatowaniu.

Inna często wykorzystywana funkcja to LEWY (ang. LEFT). Zwraca ona pierwsze znaki w ciągu tekstowym na podstawie określonej liczby znaków. Natomiast funkcja PRAWY (ang. RIGHT) zwraca ostatnie znaki w ciągu tekstowym, na podstawie określonej liczby znaków.

Do obliczenia długości tekstu służy funkcja DŁ (ang. LEN). Funkcja zwraca liczbę znaków ciągu tekstowego. Informacja o liczbie znaków w analizowanym tekście może być podstawą do odmiennego formatowania, wykonania operacji bądź być podstawą analizy.

Pozyskiwane dane tekstowe bardzo często zawierają dużą ilość zbędnych odstępów. Na etapie porządkowania danych należy zawsze bezwzględnie je usunąć. Funkcja USUŃ.ZBĘDNE.ODSTĘPY (ang. CLEAN) usuwa wszystkie spacje z tekstu, oprócz pojedynczych spacji występujących między słowami.

Dane ze źródeł zewnętrznych zwykle są także niejednorodne pod względem wielkości liter. Funkcja LITERY.MAŁE (ang. LOWER) konwertuje wszystkie duże litery w ciągu tekstowym na małe, z kolei funkcja LITERY.WIELKIE (ang. UPPER) konwertuje małe litery na wielkie.

W przypadku rozbudowanych arkuszy z dużą ilością różnorodnych informacji często istnieje potrzeba sprawdzenia wystąpienia określonego tekstu. Funkcja SZUKAJ.TEKST (ang. FIND) służy do odnajdywania jednego ciągu tekstowego

wewnątrz innego ciągu i zwracania pozycji początkowej szukanego tekstu liczony od pierwszego znaku tekstu przeszukiwanego. Dla przykładu =SZUKAJ.TEKST(?a?; ?jabłka?) zwraca wartość 2, ponieważ „a” to druga litera w wyrazie „jabłka”. W funkcji SZUKAJ.TEKST każdy znak jest zawsze liczony jako jedno wystąpienie, bez względu na ustawiony język domyślny.

Funkcja FRAGMENT.TEKSTU (ang. MID) zwraca określoną liczbę znaków z ciągu tekstowego, począwszy od wskazanej pozycji. Funkcja ta jest często stosowana do skrócenia tekstu lub łączona z funkcją SZUKAJ.TEKSTU.

Kolejnym błędem, który obniża jakość i możliwości wykorzystania danych w dalszych analizach, jest przechowywanie w jednej komórce wielu danych (informacji). W danych dotyczących sprzedaży Firmy X taka sytuacja miała miejsce w przypadku informacji o mieście i województwie zawartych w jednej komórce. Rozdzielenie tych danych do osobnych komórek pozwoliło na późniejsze analizowanie sprzedaży zarówno według miast, jak i według województw (rysunek 2.5).

	G	H	I	J	K
1		województwo WP		DS	DŚ
2		powiat	kaliski	kamiennogórski	kamiennogórski
3		miasto	Kalisz	Kamienna Góra	Kamienna Góra
4		jednostka	kaliski miasto Kalisz (WP)	kamiennogórski miasto Kamienna Góra (DS)	kamiennogórski miasto Kamienna Góra (DŚ)

Rysunek 2.5. Efekt pozyskania nazwy województwa i miasta

Źródło: Opracowanie własne.

W kolumnie *Zmienna* znajdowały się informacje nt. nazwy artykułu i jednostki, w których podana jest wartość. Było to nieczytelne i nie pozwalało na dokonanie analizy. Informacje te należało umieścić w dwóch osobnych kolumnach, wykorzystując odpowiednie funkcje tekstowe (rysunek 2.6).

	A	G	H	I
1			województwo WP	
2			powiat	kaliski
3			miasto	Kalisz
4	Data ewidencji stanu sprzedaży od ostatniej	artykuł	jednostka	kaliski miasto Kalisz (WP)
5	Fri Jan 31 21:17:22 - 2014	Gruszki świeże	zł	11896,96382
6	Fri Feb 28 11:57:35 - 2014	Gruszki świeże	zł	28453,92861
7	Mon Mar 31 10:29:59 - 2014	Gruszki świeże	zł	13612,48848
8	Wed Apr 30 06:47:24 - 2014	Gruszki świeże	zł	12609,83092
9	Sat May 31 00:25:48 - 2014	Gruszki świeże	zł	12031,14989

Rysunek 2.6. Efekt pozyskania nazwy artykułu i jednostki

Źródło: Opracowanie własne.

W kolumnie *Data* znajdowały się informacje o datach sprzedaży. Takie skumulowane informacje nie pozwalały na żadne operacje analityczne, nie pozwalały na wnioskowanie po dacie. W trzech kolumnach – dzień, miesiąc, rok – należało zamieścić te informacje (rysunek 2.7). W tym celu należało wykorzystać odpowiednie funkcje tekstowe.

	B	C	D	F	G	H	
1					województwo	WP	DS
2					powiat	kaliski	kamiennog
3					miasto	Kalisz	Kamienna C
4	dzień	miesiąc	rok	artykuł	jednostka	kaliski miasto Kalisz (WP)	kamiennog
5	31	Jan	2014	Gruszki świeże	zł	11896,96382	
6	28	Feb	2014	Gruszki świeże	zł	28453,92861	
7	31	Mar	2014	Gruszki świeże	zł	13612,48848	
8	30	Apr	2014	Gruszki świeże	zł	12609,83092	
9	31	May	2014	Gruszki świeże	zł	12031,14989	
10	30	Jun	2014	Gruszki świeże	zł	9224,673598	
11	31	Jul	2014	Gruszki świeże	zł	12874,53771	
12	31	Aug	2014	Gruszki świeże	zł	20535,89686	
13	30	Sep	2014	Gruszki świeże	zł	23969,07868	

Rysunek 2.7. Efekt podziału tekstu na części daty

Źródło: Opracowanie własne.

Podstawowe formuły wykorzystane w ramach niniejszego studium przypadku zaprezentowano w tabeli 2.2.

Tabela 2.2. Formuły tekstowe wykorzystane w celu przygotowania danych sprzedaży do analizy

Nazwa kolumny	Formuła
Województwo	=FRAGMENT.TEKSTU(C2;SZUKAJ.TEKST("(?";C2)+1;2)
Powiat	=FRAGMENT.TEKSTU(D4;1;SZUKAJ.TEKST(" ";D4))
Miasto	=FRAGMENT.TEKSTU(D4;SZUKAJ.TEKST("miasto";D4)+7;SZUKAJ.TEKST("";D4)-SZUKAJ.TEKST("miasto";D4)-8)
Artykuł	=FRAGMENT.TEKSTU(B3;1; SZUKAJ.TEKST(? sprzed?;B3)-1)
Jednostka	=FRAGMENT.TEKSTU(B3; SZUKAJ.TEKST(?w?;B3;DŁ(B3)-5)+2;DŁ(B3))
Dzień	=FRAGMENT.TEKSTU(A3;9;2)
Miesiąc	=FRAGMENT.TEKSTU(A3;5;3)
Rok	=PRAWY(A3;4)

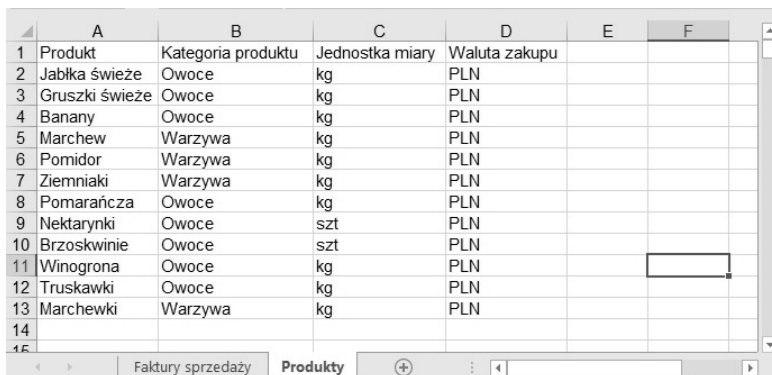
Źródło: Opracowanie własne.

Dodatkowo w nowo utworzonej kolumnie utworzono formułę, która połączyła dzień, miesiąc, rok, pozyskując w ten sposób kompletną datę. W tym celu wykorzystano funkcję ZŁĄCZ.TEKSTY. Wyniki przygotowania danych dotyczących sprzedaży do analizy zawiera plik *r2_dane_sprzedazy_o.xlsx*.

2.3. Przypadek 3 – wykorzystanie narzędzi danych w celu poprawy ich jakości

Pracownicy Firmy X w swojej codziennej pracy wykorzystują arkusz kalkulacyjny MS Excel do analizy danych. Informacje na temat faktur pozyskiwane są z różnych systemów transakcyjnych i umieszczane w arkuszu (plik *r1_faktury_sprzedazy_no.xlsx*). Plik arkusza kalkulacyjnego staje się bardzo duży, a liczba wierszy powoduje brak czytelności. Zaproponowano proces mający na celu eliminację powtarzających się danych. Główna idea polegała na trzymaniu danych w różnych arkuszach, a w razie potrzeby linkowania do danych. Taki sposób przechowywania danych zwiększa ich bezpieczeństwo i zmniejsza ryzyko powstania niespójności (w szczególności problemów anomalii). Dla zwiększenia czytelności danych stworzono dodatkowe arkusze, zawierające unikalne dane. Pomogło to w późniejszej analizie danych, a ponadto było podstawą do utworzenia odpowiednich reguł poprawności. W realizacji wykorzystano możliwości MS Excel zawarte w grupie *Narzędzia danych* (wstążka *Dane*).

Zauważono, że informacje na temat nazwy produktu, kategorii, jednostki miary i waluty zakupu przy każdym produkcie się powtarzają. Są one bezpośrednio powiązane z nazwą produktu. Utworzono zatem dodatkowy arkusz o nazwie *Produkty* i skopiowano do niego zawartość kolumn: *Produkt*, *Kategoria produktu*, *Jednostka miary* i *Waluta zakupu*. Ponieważ dane się powtarzały, usunięto duplikaty za pomocą polecenia *Usuń Duplikaty* z grupy *Narzędzia danych*. Efekt obrazuje rysunek 2.8.

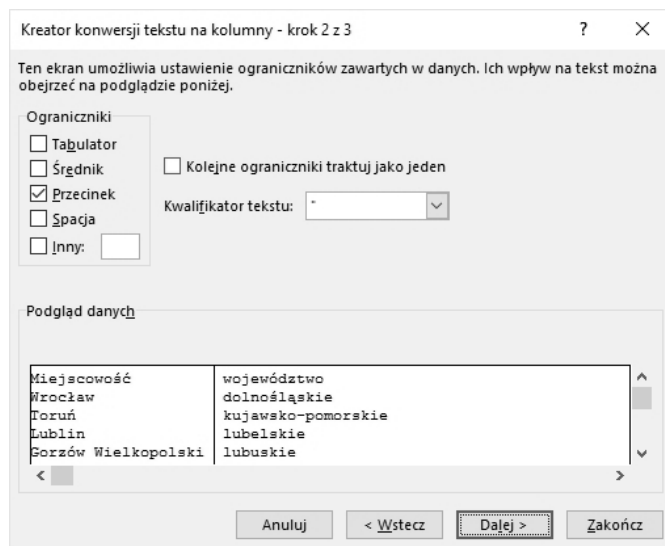


	A	B	C	D	E	F
1	Produkt	Kategoria produktu	Jednostka miary	Waluta zakupu		
2	Jabłka świeże	Owoce	kg	PLN		
3	Gruszki świeże	Owoce	kg	PLN		
4	Banany	Owoce	kg	PLN		
5	Marchew	Warzywa	kg	PLN		
6	Pomidor	Warzywa	kg	PLN		
7	Ziemniaki	Warzywa	kg	PLN		
8	Pomarańcza	Owoce	kg	PLN		
9	Nektarynki	Owoce	szt	PLN		
10	Brzoskwinie	Owoce	szt	PLN		
11	Winogrona	Owoce	kg	PLN		
12	Truskawki	Owoce	kg	PLN		
13	Marchewki	Warzywa	kg	PLN		
14						
15						

Rysunek 2.8. Uporządkowana tabela z danymi o produktach (arkusz *Produkty*)

Źródło: Opracowanie własne.

Dane na temat miejscowości i województwa zapisane są w jednej kolumnie. Jest to błędne z punktu widzenia późniejszej analizy danych. Dokonano zatem podziału tej kolumny na dwie osobne, wykorzystując polecenie *Tekst jako kolumny* z grupy *Narzędzia danych* (rysunek 2.9).



Rysunek 2.9. Przykład podziału tekstu na dwie kolumny

Źródło: Opracowanie własne.

Ponadto za pomocą formuły `=Z.WIELKIEJ.LITERY(N1)` zamieniono pierwsze litery w nazwie województwa na duże.

Dodatkowo w arkuszu głównym niepotrzebnie powtarzane były informacje na temat klientów. Zauważono, iż *Miejscowość* i *Województwo* bezpośrednio uzależnione są od nazwy klienta. Stworzono kolejny arkusz o nazwie *Klienci* (rysunek 2.10). W arkuszu tym znalazły się skopiowane kolumny: *Klient*, *Miejscowość* i *Województwa*. Aby pozbyć się powtarzających się danych, usunięto duplikaty.

	A	B	C
1	Klient	Miejscowość	Województwo
2	ABC Sp. z o.o.	Wrocław	Dolnośląskie
3	Owoce i warzywa Sp. z o.o.	Toruń	Kujawsko-Pomorskie
4	Hurtownia spożywcza	Lublin	Lubelskie
5	Jan Kowalski, Sp. J.	Gorzów Wielkopolski	Lubuskie
6	Żywność hurt i detal Sp. z o.o.	Łódź	Łódzkie
7	Zakłady spożywcze SA	Kraków	Małopolskie
8	Zdrowa żywność SA	Warszawa	Mazowieckie
9	Wegan Hurt Sp z o.o.	Opole	Opolskie
10	Hurtownia dla wegetarian sp z o.o.	Rzeszów	Podkarpackie
11	Hurtownia spożywcza w Białymstoku	Białystok	Podlaskie
12	Żywność Gdańska Sp. z o.o.	Gdańsk	Pomorskie
13	Multiwybór owoców Sp. z o.o.	Katowice	Śląskie
14	Adam Nowak Sp.J.	Kielce	Świętokrzyskie
15	Mazurskie wyroby Sp. z o.o.	Olsztyn	Warmińsko-Mazurskie
16	Najlepsza żywność Sp. z o.o.	Poznań	Wielkopolskie
17	Owoce z Zachodniego Pomorza, Sp. z o.o.	Szczecin	Zachodniopomorskie

Rysunek 2.10. Uporządkowana tabela z danymi klientów (arkusz *Klienci*)

Źródło: Opracowanie własne.

Kolejnym arkuszem, jaki wyodrębniono, jest arkusz *Faktury_sprzedazy* zawierający następujące informacje: numer faktury, datę wystawienia, nazwę sprzedawcy, nazwę klienta, termin płatności, termin zapłaty oraz sposób zapłaty (rysunek 2.11). Przed przekazaniem danych do analiz usunięto duplikaty.

	A	B	C	D	E	F	G
1	Nr dokumentu sprzedaży	Data	Sprzedawca	Klient	Termin płatności	Termin zapłaty	Sposób zapłaty
2	FA846532/64/646	07.01.2014	Targ owocowy w Polsce, sieć hurtowni	ABC Sp. z o.o.	21.01.2014	04.02.2014	przelew
3	FA847932/64/646	07.01.2014	Targ owocowy w Polsce, sieć hurtowni	Owoce i warzywa Sp. z o.o.	21.01.2014	25.01.2014	przelew
4	FA777212/64/646	07.01.2014	Międzynarodowy targ owocowy	Hurtownia spożywcza	21.01.2014	28.01.2014	przelew
5	FA8074112/64/646	07.01.2014	Pośrednictwo sprzedaży warzyw i owoców SA	Jan Kowalski, Sp. J.	21.01.2014	02.02.2014	przelew
6	FA8017532/64/646	07.01.2014	Pośrednictwo sprzedaży warzyw i owoców SA	Zywność hurt i detal Sp. z o.o.	21.01.2014	31.01.2014	gotówka
7	FA8490122/64/646	07.01.2014	Targ rolny	Zakłady spożywcze SA	21.01.2014	26.01.2014	gotówka
8	FA8090156/64/647	08.01.2014	Pośrednictwo sprzedaży warzyw i owoców SA	Zdrowa żywność SA	22.01.2014	22.01.2014	przelew
9	FA808732/64/647	08.01.2014	Pośrednictwo sprzedaży warzyw i owoców SA	Wegan Hurt Sp z o.o.	22.01.2014	27.01.2014	gotówka
10	FA807232/64/647	08.01.2014	Pośrednictwo sprzedaży warzyw i owoców SA	Hurtownia dla wegetarian sp z o.o.	22.01.2014	30.01.2014	gotówka
11	FA8072115/64/647	08.01.2014	Pośrednictwo sprzedaży warzyw i owoców SA	Hurtownia spożywcza w Białymstoku	22.01.2014	31.01.2014	gotówka
12	FA8017532/64/647	08.01.2014	Pośrednictwo sprzedaży warzyw i owoców SA	Zywność Gsańska Sp. z o.o.	22.01.2014	01.02.2014	gotówka
13	FA847732/64/647	08.01.2014	Targ owocowy w Polsce, sieć hurtowni	Multywybór owoców Sp. z o.o.	22.01.2014	06.02.2014	gotówka

Rysunek 2.11. Uporządkowana tabela z informacjami o fakturach (arkusz *Faktury_sprzedazy*)

Źródło: Opracowanie własne.

Ostatnią grupą informacji istotną dla późniejszej analizy sprzedaży były dane dotyczące sprzedanych produktów w ramach poszczególnych faktur. Na fakturach znajdowały się następujące dane sprzedanych pozycji: nazwa produktu, cena sprzedaży netto, rabat, ilość i kwota zapłaty (rysunek 2.12). Zostały one umieszczone w arkuszu *Pozycje_faktur*.

	A	B	C	D	E	F
1	Nr dokumentu sprzedaży	Produkt	Cena sprzedaży netto	Rabat	Ilość	Kwota zapłaty
2	FA846532/64/646	Jabłka świeże	2,00 zł	5%	115	264,50 zł
3	FA847932/64/646	Gruszki świeże	4,00 zł	10%	184	115,00 zł
4	FA777212/64/646	Banany	3,40 zł	0%	20	74,29 zł
5	FA8074112/64/646	Marchew	1,50 zł	5%	109	208,73 zł
6	FA8017532/64/646	Pomidor	5,00 zł	0%	57	172,50 zł
7	FA8490122/64/646	Ziemniaki	3,00 zł	0%	65	607,20 zł

Rysunek 2.12. Uporządkowana tabela z informacjami o pozycjach na fakturach (arkusz *Pozycje_faktur*)

Źródło: Opracowanie własne.

W Firmie X faktury bardzo często wprowadzane są ręcznie przez wyznaczone do tego osoby. Ręczne wprowadzanie danych może prowadzić do błędów. Aby zapewnić wysoki poziom jakości danych, firma zdecydowała się na wykorzystanie narzędzia *Sprawdzanie poprawności danych* (wstążka *Dane > Poprawność danych*). Za pomocą tego narzędzia można kontrolować typ danych lub wartości wprowadzane przez użytkowników w komórce. Można na przykład ograniczyć wprowadzanie danych do określonego zakresu dat, ograniczyć wybory przy użyciu listy lub zagwarantować, że są wprowadzane tylko całkowite liczby dodatnie [Walkenbach, 2013].

Sprawdzanie poprawności danych to funkcja programu MS Excel, którą bardzo często wykorzystuje się w firmach. Za pomocą niej można definiować ograniczenia określające, jakie dane mogą lub powinny zostać wprowadzone w komórce.

Sprawdzanie poprawności danych można skonfigurować, aby uniemożliwić użytkownikom wprowadzanie nieprawidłowych danych. Można też zezwalać użytkownikom na wprowadzanie nieprawidłowych danych i ustawić ostrzeżenia, gdy próbują wpisać je w komórce. Ponadto można udostępnić komunikaty definiujące dane wejściowe oczekiwane dla komórki oraz instrukcje ułatwiające użytkownikom poprawianie ewentualnych błędów.

W przypadku danych dotyczących sprzedaży w Firmie X stworzono odpowiednie reguły poprawności dla nowo wprowadzanych faktur i określono, że:

- *Produkt* to pozycja z arkusza *Produkt*,
- *Kategoria produktu* to *Owoce* bądź *Warzywa*,
- *Cena sprzedaży netto* jest większa od zera,
- *Cena sprzedaży netto* jest wyższa niż *Cena zakupu*,
- *Rabat* jest większy lub równy zero i mniejszy niż 20% itd.

Dla powyższych reguł wpisano odpowiednie kryteria poprawności, komunikat wejścia, jak i komunikat błędu.

Wyniki przygotowania danych sprzedaży do analizy zawiera plik *r2_faktury_sprzedazy_o.xlsx*.

2.4. Możliwe zastosowania MS Excel w biznesie w formie zadań

W celu zobrazowania możliwości wykorzystania narzędzia MS Excel w zakresie przygotowania danych do analizy, można posłużyć się poniższymi zadaniami:

1. Firma X przygotowuje do analizy tabelę z informacjami o danych giełdowych, które pozyskała w ramach realizacji czwartego studium przypadku w rozdziale 1 (plik *r1_dane_gieldowe.xlsx*). Firma musi usunąć zbędne kolumny zawierające takie same informacje lub niezawierające żadnych danych. Ponadto firma musi zmienić format odpowiednich kolumn na walutowy.
2. Firma X przygotowuje do analizy tabelę z informacjami o osobach z wyższym wykształceniem (plik *r1_wyzsze_wyksztalzenie.xlsx*) i wynagrodzeniach brutto (plik *r1_wynagrodzenia_brutto.xlsx*), które pozyskała w ramach piątego studium przypadku w rozdziale 1. W tym celu musi usunąć wszystkie zbędne wiersze opisujące tabele z danymi.

Zakończenie

Niniejszy rozdział wskazuje na bardzo duże możliwości wykorzystania narzędzia MS Excel 2016 na etapie przygotowywania danych do analizy. Pełne spektrum

narzędzi służących do formatowania danych w połączeniu z łatwością dodawania i usuwania kolumn/wierszy daje wręcz nieograniczone możliwości przetwarzania danych, tak aby były one poprawnie przygotowane do analizy oraz czytelne i estetyczne. Umiejętne wykorzystanie prezentowanych mechanizmów jest niezbędną umiejętnością, którą powinien posiadać każdy analityk biznesowy [Ren, 2014].

Powyższe rezultaty jednoznacznie wskazują, że MS Excel jest bardzo użytecznym narzędziem dla zaawansowanej analizy danych. Ważne jest postrzeganie analizy danych jako całościowy, wieloetapowy proces [Winston, 2014]. W pierwszym kroku należy zadbać o pozyskanie wysokiej jakości danych i przeniesienie ich do MS Excel (opisano ten etap w rozdziale 1). Dane te najczęściej nie nadają się od razu do analizy. Należy zatem je oczyścić i odpowiednio sformatować, co przedstawiono w niniejszym rozdziale. Dopiero poprawnie sformatowane dane są podstawą do dalszych działań analitycznych. Dane o niskiej jakości, jak i dane niepoprawnie sformatowane, powodować będą błędy w analityce biznesowej. Nieczytelna prezentacja danych także będzie skutkowałą najczęściej niepoprawną interpretacją danych. Kolejne rozdziały prezentują mechanizmy analizy danych.

Bibliografia

1. Alexander M., Decker J. (2014), *Microsoft Business Intelligence Tools for Excel Analysts*, Wiley.
2. European Institute of Public Administration (EIPA) (2013), *The Common Assessment Framework (CAF). Improving Public Organisations through Self-Assessment*, CAF.
3. Harper E. (2015), *ISO 8000 44 Success Secrets – 44 Most Asked Questions On ISO 8000 – What You Need To Know*, Emereo Publishing.
4. ISO/TS 8000-1:2011, Data quality – Part 1: Overview, 2011.
5. McGilvray D. (2008), *Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information (TM)*, Morgan Kaufmann.
6. Nussbaumer Knaflig C. (2015), *Storytelling with Data: A Data Visualization Guide for Business Professionals*, Wiley.
7. Redman T.C. (2001), *Data Quality. TheField Guide*.
8. Ren J. (2014), *Excel for Business Professionals and Managers*, Amazon Digital Services, Inc.
9. Walkenbach J. (2013), *Excel 2013 Formulas*, Wiley.
10. Walkenbach J., (2013), *Excel 2013 PL. Biblia*, Helion, Gliwice.
11. Winston W. (2014), *Microsoft Excel 2013 Data Analysis and Business Modeling*, Microsoft Press.
12. Wrycza S. (red.) (2010), *Informatyka ekonomiczna. Podręcznik akademicki*, PWE, Warszawa.

Rozdział 3

Analiza danych z wykorzystaniem zaawansowanych formuł i funkcji

Michał Kuciapski

Wstęp

Kluczowe dla dokonywania analizy danych biznesowych zawartych w arkuszu kalkulacyjnym jest stosowanie odpowiednich formuł i funkcji [McFedries, 2013]. Ze względu na dużą liczbę dostępnych funkcji podzielone one zostały na kilkanaście kategorii, jak: finansowe, daty i godziny, matematyczne, statystyczne, wyszukiwania i adresu, bazy danych, tekstowe, logiczne, informacyjne, inżynierskie, modułowe, zgodność oraz sieci web. Szczególnie istotne dla analizy i wnioskowania są funkcje ujęte w kategoriach: logiczne, matematyczne, statystyczne, wyszukiwania i adresu oraz bazodanowe. Łączenie formuł i funkcji pozwala przetworzyć w złożony sposób dane dla uzyskania wyników umożliwiających interpretację zjawisk ekonomicznych [Coontz, 2001], powiązanych z takimi procesami, jak: sprzedaż, logistyka, finanse czy kontroling. Formuły w arkuszu kalkulacyjnym pozwalają wykonywać operacje na danych dla przetworzenia ich w użyteczne informacje, jak np. EBITDA (ang. *Earnings Before Interest, Taxes, Depreciation and Amortization*) – zysk przedsiębiorstwa przed potrąceniem odsetek od zaciągniętych kredytów, podatków oraz amortyzacji.

W związku z tym, że wiele przedsiębiorstw i instytucji wykorzystuje narzędzie MS Excel dla analizy danych i wnioskowania na ich podstawie, celem niniejszego rozdziału jest przegląd oraz prezentacja możliwości wykorzystania dla przetwarzania danych biznesowych funkcji: logicznych, matematycznych, statystycznych, wyszukiwania i adresu, bazodanowych oraz finansowych. W tym zakresie zaproponowano rozwiązania pozwalające na sprawną analizę danych i interpretację wyników, w postaci studiów przypadków. Na wstępie rozdziału, w ramach pierwszego studium przypadku, przedstawiono, jak przeprowadzić analizę danych sprzedaży z wykorzystaniem funkcji matematycznych i statystycznych. W kolejnym studium przypadku zaprezentowano mechanizmy automatyzacji interpretacji wyników za pomocą formuł logicznych. W trzecim studium przypadku ukazano zastosowanie funkcji wyszukiwania i adresu dla analizy danych, na przykładzie eksplikacji struktury wykształcenia. W ramach czwartego studium przypadku poszerzono wcześniejszą analizę danych sprzedaży, wskazując, jak uzyskać dane

przekrojowe, dzięki zastosowaniu funkcji bazodanowych. Ostatnie studium przypadku prezentuje proces określenia rentowności inwestycji za pomocą funkcji finansowych. Rozdział kończy wskazanie zastosowań arkusza MS Excel w biznesie oraz podsumowanie.

3.1. Przypadek 1 – analiza danych sprzedaży z wykorzystaniem funkcji matematycznych i statystycznych

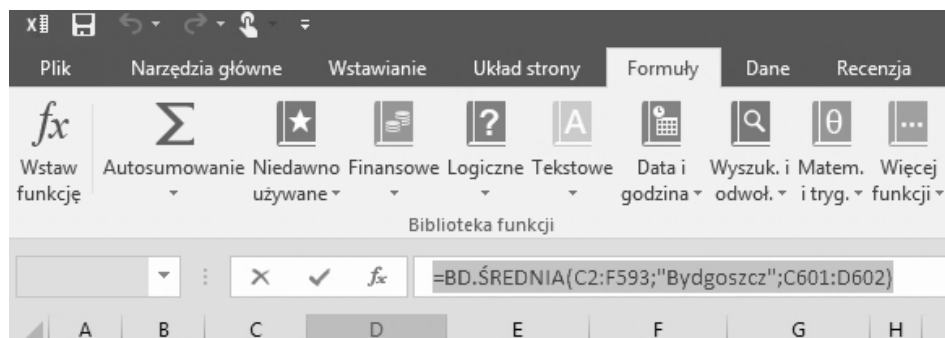
Niezbędną informacją dla firmy pragnącej wejść na dany rynek jest znajomość jego wartości oraz poziomu nasycenia. Pozwala to określić zarówno poziom ryzyka wprowadzenia oferty na danym rynku, jak i wybrać potencjalnie najlepszy spośród analizowanych [Penza, Bansa, 2001].

Niniejsze studium przypadku dotyczy Firmy X, która działa w kilkunastu największych miastach w Polsce. Przedsiębiorstwo posiadało w tym zakresie kilkanaście hurtowni owoców, przede wszystkim jabłek i gruszek, w całym kraju i planowało otworzyć kolejny oddział, w mieście o kilkudziesięciu tysiącach mieszkańców. Pod uwagę brane były dwa miasta: Kartuzy i Kwidzyn. Dla podjęcia decyzji konieczne było uzyskanie szeregu informacji dotyczących sprzedaży owoców w odniesieniu do obu miast:

- średnia, minimalna i maksymalna sprzedaż jabłek i gruszek dla porównania wielkości rynków;
- wariancja dla określenia poziomu zróżnicowania sprzedaży dla poszczególnych miesięcy;
- odchylenie standardowe dla wskazania przedziału, w którym znajduje się większość wartości sprzedaży dla poszczególnych miesięcy;
- współczynnik zmienności, pozwalający w zestandaryzowany sposób określić poziom zróżnicowania sprzedaży w poszczególnych miesiącach.

Dane do analizy zawiera arkusz *dane sprzedaży* w skoroszycie *r3_analiza_sprzedazy_formuly.xlsx*. Arkusz został skopiowany z pliku *r2_dane_sprzedazy_o.xlsx*, który został odpowiednio przygotowany do analizy w ramach przypadku 2 w rozdziale 2. W celu uzyskania potrzebnych danych biznesowych zastosowano operatory matematyczne w połączeniu z funkcjami MS Excel. Ważne było zastosowanie dla definiowanych formuł wyrażeń opartych o adresy komórek. Zapewniło to stabilność formuł oraz ich dostosowywanie się do miejsca wstawienia w arkuszu, dzięki stosowaniu adresowania względnego [McFedries, 2013]. W ten sposób ograniczono czas na poprawianie formuł w przypadku zmian w arkuszu, a w szczególności dodawania lub usuwania nowych wierszy i kolumn.

Wyrażenia wprowadzono bezpośrednio w pasku formuły, dostępnym poprzez wstążkę *Formuły* w arkuszu kalkulacyjnym MS Excel, co prezentuje rysunek 3.1.



Rysunek 3.1. Wstążka *Formuły* dla wprowadzania wyrażeń przetwarzania danych

Źródło: Opracowanie własne.

W wybranych przypadkach, gdzie przetwarzanie danych jest wysoce złożone, dostępne funkcje są niewystarczające. Na przykład nie jest możliwe wyliczenie zmiany wartości kapitału dla poszczególnych lat przez rentiera z uwzględnieniem automatycznego pobierania inflacji i WIBORU (ang. *Warsaw Interbank Offered Rate*) na zasadzie parowania stron. W takiej sytuacji należy opracować własne funkcje, korzystając w tym celu z dedykowanego języka programowania dla narzędzi Office – VBA (*Visual Basic for Applications*) [Jackson, Staunton, 2006]. W prezentowanym studium przypadku dostępne w MS Excel funkcje były wystarczające dla wyliczenia założonych wskaźników.

Przed przystąpieniem do wprowadzania formuł i funkcji dla analizy danych przeprowadzono weryfikację poprawności formatu danych. Jeśli okazałoby się, że dane liczbowe były traktowane przez MS Excel jako tekst, nie byłoby możliwe wykonywanie jakichkolwiek formuł bazujących na liczbach, takich jak statystyczne czy matematyczne. Zastosowano sposoby konwersji danych zawarte w rozdziale 2 niniejszego opracowania dla uzyskania wartości liczbowych dla lat oraz miesięcy.

Dla uzyskania pożądanego danych analitycznych wykorzystano funkcje matematyczne i statystyczne. W tym zakresie MS Excel posiada kilkadziesiąt funkcji. Funkcje matematyczne, wbrew nazwie, pozwalają nie tylko na wykonywanie operacji matematycznych, ale także na powiązanie ich z wyrażeniami warunkowymi. Taką funkcją jest SUMA.JEŻELI. Pozwala ona przetwarzać dane, uprzednio filtrując je, w ten sposób zawężając badane przedziały [Abramowitz, Stegun, 1964]. Przydatne funkcje matematyczne zestawiono w tabeli 3.1.

Tabela 3.1. Kluczowe funkcje matematyczne

Funkcja	Znaczenie	Składnia
SUMA. JEŻELI	Zsumowanie wartości z komórek określonych przez podane kryterium	SUMA.JEŻELI(zakres; kryteria; [suma_zakres])
SUMA. WARUNKÓW	Zsumowanie wszystkich argumentów, które spełniają określone kryteria	SUMA.WARUNKÓW(suma_zakres; kryteria_zakres1;kryteria1;[kryteria_zakres2;kryteria2];...)
SUMY. CZĘŚCIOWE	Zwrócenie sumy częściowej na liście lub w bazie danych, zgodnie z kryteriami filtrowania i grupowania	SUMY.CZĘŚCIOWE(funkcja_nr; odw1;[odw2];...)

Źródło: Opracowanie własne.

Uniwersalność wykorzystanych funkcji zapewniana jest poprzez ich argumenty, pozwalające wskazać zarówno dane wejściowe, jak i w wybranych przypadkach określić ich działanie. W związku z tym wiele argumentów ma charakter opcjonalny, czyli mogą, ale nie muszą być specyfikowane. W tabelach przeglądu funkcji, jak tabela 3.1, parametry tego typu są podawane w nawiasach kwadratowych.

Istotnym dopełnieniem funkcji matematycznych, ważnym w procesie analizy danych opisujących zjawiska, są funkcje statystyczne. Analogicznie do funkcji matematycznych wybrane z nich pozwalają filtrować dane z uwzględnieniem filtrowania. Przykładami są funkcje ŚREDNIA.JEŻELI czy LICZ.JEŻELI. Szczególnie użyteczne, w tym te, które miały zastosowanie w prezentowanym studium przypadku, zestawiono w tabeli 3.2. Zauważalne jest wzajemne uzupełnianie się funkcji matematycznych i statystycznych dla wyliczania wartości przydatnych dla wnioskowania.

Tabela 3.2. Wybrane funkcje statystyczne

Funkcja	Znaczenie	Składnia
ODCH.STAND. POPUL	Obliczenie odchylenia standardowego dla całej populacji, ze zignorowaniem wartości logicznych i tekstu	ODCH.STAND.POPUL(liczba1;[liczba2];...)
WARIANCJA. POPUL	Obliczenie wariancji na podstawie całej populacji	WARIANCJA.POPUL(liczba1;[liczba2];...)
LICZ.JEŻELI	Zliczenie liczby komórek, które spełniają dane kryteria	LICZ.JEŻELI(zakres; kryteria)
ŚREDNIA. JEŻELI	Zwrócenie średniej arytmetycznej ze wszystkich komórek z zakresu, które spełniają podane kryteria	ŚREDNIA.JEŻELI(zakres; kryteria;[średnia_zakres])
LICZ. WARUNKI	Zliczenie, ile razy zdefiniowane kryteria zostały spełnione w wielu zakresach	LICZ.WARUNKI(zakres_kryterium1;kryterium1; [zakres_kryterium2;kryterium2]...)
WYST. NAJCZĘŚCIEJ	Zwrócenie wartości najczęściej występującej w tablicy albo w zakresie danych.	WYST.NAJCZĘŚCIEJ.TABL(liczba1;[liczba2];...) WYST.NAJCZĘŚCIEJ.WART(liczba1;[liczba2];...)

Źródło: Opracowanie własne.

Arkusz z danymi, czyli listą produktów w zakresie warzyw i owoców, został przygotowany poprzez ukrycie wszystkich zbędnych kolumn (polecenie *Ukryj*) – oprócz tych dla Kartuz i Kwidzyna. Ponadto dzięki włączeniu autofiltrowania (wstążka *Dane*, polecenie *Filtruj*), szerzej zaprezentowanego w rozdziale 5, uzyskano wartości sprzedaży wyłącznie jabłek i gruszek. Tak opracowane dane zostały skopiowane do osobnego arkusza, aby ukryte poprzez filtrowanie wiersze nie były brane pod uwagę. W kolejnym kroku wprowadzono stosowne formuły dla obu miast:

- średnia (pierwsza komórka zakresu : ostatnia komórka zakresu),
- min (pierwsza komórka zakresu : ostatnia komórka zakresu),
- max (pierwsza komórka zakresu : ostatnia komórka zakresu),
- wariancja.popol (pierwsza komórka zakresu : ostatnia komórka zakresu),
- odch.stand.popol (pierwsza komórka zakresu : ostatnia komórka zakresu),
- współczynnik zmienności, wynoszący: odchylenie standardowe / średnia * 100%.

Kopiowanie formuł ze względu na stosowanie adresowania względnego dokonuje przesunięcia zakresu podawanego w funkcjach. MS Excel wyłapuje, że jest to potencjalnie niepożądane i poprzez operację *Aktualizuj formułę* pozwala, aby zakresy w trakcie kopiowania formuły nie ulegały zmianie. W przypadku adresowania względnego podczas kopiowania formuły dokonywane jest przekształcenie formuły zgodnie z relacją pomiędzy komórkami. W związku z tym, że niepożądane było, aby formuła zmieniała się, w takiej sytuacji zastosowano adresowanie bezwzględne, poprzedzając adresy komórek, czyli identyfikatory wierszy i kolumn, symbolem \$. Uzyskane wartości dla poszczególnych formuł przedstawia rysunek 3.2.

=ODCH.STAND.POPUL(F3:F400)										
	A	B	C	D	E	F	G	H	I	J
1						KP	KP			
2	dzień	miesiąc	rok	artykuł	jednostka	Kartuzy	Kwidzyn	Parametr	Kartuzy	Kwidzyn
3	30	Kwi	1996	Jabłka świeże	zł	23100,19126	28511,40414	średnia	15 251,82 zł	15 834,30 zł
4	30	Cze	1996	Jabłka świeże	zł	9579,156216	14804,93428	min	8,16 zł	101,96 zł
5	31	Lip	1996	Jabłka świeże	zł	11234,36205	7389,87732	max	29 968,88 zł	29 951,07 zł
6	29	Lut	1996	Jabłka świeże	zł	25071,52852	18326,50241	wariancja	75046414	74154689
7	31	Maj	1996	Jabłka świeże	zł	17096,26753	11871,47242	odchylenie standardowe	8 662,93 zł	8 611,31 zł
8	31	Mar	1996	Jabłka świeże	zł	3792,604157	28337,78832	współczynnik zmienności	57%	54%

Rysunek 3.2. Analiza sprzedaży w oparciu o funkcje matematyczne i statystyczne

Źródło: Opracowanie własne.

Zgodnie z wynikami prezentowanymi na rysunku 3.2 wartości sprzedaży dla obu analizowanych miast były bardzo zbliżone. Niemniej średnia wartość sprzedaży gruszek świeżych i jabłek świeżych dla poszczególnych miesięcy (data na koniec miesiąca) była nieznacznie większa dla Kwidzyna (15 834 zł względem 15 252 zł). Analogicznie wartość minimalna sprzedaży dla pojedynczego miesiąca

była wyższa dla Kwidzyna (102 zł w porównaniu do 8 zł). Odchylenie standardowe było wyższe dla Kartuz w porównaniu do Kwidzyna – 8663 zł względem 8611 zł. Wartość odchylenia standardowego służy weryfikacji zmienności wartości sprzedaży w poszczególnych miesiącach i może być stosowana wyłącznie ze względu na bardzo zbliżoną średnią wartość sprzedaży dla obu miast. Dla wnioskowania powinno stosować się jednak współczynnik zmienności [Rao, 2009]. Zgodnie z nim zmienność była mniejsza dla Kwidzyna (54%) względem Kartuz (57%). Oznacza to, że zmiany poziomu sprzedaży były mniejsze w Kwidzynie, co jest na pewno korzystne dla firmy. Zauważalna była, analogicznie do pozostałych miar, mała różnica pomiędzy miastami, jak i sama wysoka wartość wskaźnika. W obu zatem miastach występowały znaczne zmiany w poziomie sprzedaży. Obliczone miary wskazały na wysokie podobieństwo obu rynków, przez co trudność w jednoznacznym określeniu, który z nich jest korzystniejszy z punktu widzenia poziomu i stabilności sprzedaży owoców – jabłek i gruszek. Wyniki analiz zrealizowanych w ramach niniejszego studium przypadku zawiera arkusz *analiza sprzedazy_1* w pliku *r3_analiza_sprzedazy_formuly.xlsx*.

3.2 Przypadek 2 – analiza danych sprzedaży z zastosowaniem funkcji logicznych

Stosowanie formuł oraz funkcji pozwala na analizę danych w wielu sytuacjach. Często jednak sposób wyliczenia wartości, takich jak wynagrodzenie pracownika czy łączny koszt inwestycji, ma różną formułę. W przypadku określenia wynagrodzeń pracowników dla poszczególnych z nich może to być:

- pensja plus dodatek stażowy, w przypadku umowy o pracę;
- liczba godzin razy stawka za godzinę, w sytuacji umowy zlecenia;
- liczba wyprodukowanych sztuk razy wynagrodzenie za sztukę, w przypadku pracy na akord.

Dla dostosowania formuły w zależności od sytuacji służą głównie wyrażenia i funkcje logiczne. Oprócz adaptacji formuły w zależności od danych wejściowych, funkcje logiczne są wysoce przydatne do interpretacji danych [Gamut, 1991]. W ramach pierwszego studium przypadku Firma X działająca w kilku miastach w Polsce dokonywała wyboru lokalizacji dla otworzenia nowego oddziału sprzedaży owoców i warzyw. Dla podjęcia decyzji wyliczane było m.in. odchylenie standardowe, które określa, jak duże jest zróżnicowanie sprzedaży dla poszczególnych miesięcy. W celu porównania wyników poprzez stosowaną formułę określono współczynnik zmienności (arkusz *analiza sprzedazy* w pliku *r3_analiza_sprzedazy_formuly.xlsx*). Wynik miał charakter liczbowy i problemem była konieczność znajomości przedziałów zmienności dla interpretacji jego war-

tości, czego nie było w stanie wykonać wielu z pracowników. W związku z tym przydatne było zastosowanie funkcji logicznych dla opracowania automatycznej interpretacji wyników współczynnika zmienności.

Warto pamiętać, że użyteczne funkcje logiczne są dostępne nie tylko w ramach kategorii funkcji logicznych, ale również matematycznych, statystycznych czy bazy danych. Zestawienie kluczowych funkcji z kategorii logiczne dostępnych w MS Excel zawiera tabela 3.3. Funkcje logiczne związane z kategorią bazy danych przedstawiono w ramach studium przypadku czwartego.

Tabela 3.3. Kluczowe funkcje logiczne

Funkcja	Znaczenie	Składnia
JEŻELI	Zwraca jedną wartość, jeśli podany warunek (test logiczny) ma wartość PRAWDA, albo inną wartość, jeśli ten warunek ma wartość FAŁSZ.	JEŻELI(test_logiczny; [wartość_jeżeli_prawda]; [wartość_jeżeli_fałsz])
ORAZ	Iloczyn logiczny: wynikiem funkcji jest PRAWDA, jeśli wszystkie jej argumenty mają wartość PRAWDA, lub FAŁSZ, jeśli choć jeden z jej argumentów ma wartość FAŁSZ.	ORAZ(wartość_logiczna1; [wartość_logiczna2];...)
LUB	Alternatywa logiczna: zwraca wartość PRAWDA, jeśli co najmniej jeden z argumentów ma wartość PRAWDA.	LUB(wartość_logiczna1; [wartość_logiczna2]; ...)

Źródło: Opracowanie własne.

W analizowanym przypadku użyteczna była funkcja JEŻELI. Wyróżniono pięć przedziałów, które powinna obejmować interpretacja współczynnika zmienności (V) [Cieciura, Zacharski, 2011]:

- $V < 20\%$ – mała zmienność,
- $20\% < V < 40\%$ – przeciętna zmienność,
- $40\% < V < 100\%$ – duża zmienność,
- $100\% < V < 150\%$ – bardzo duża zmienność,
- $V > 150\%$ – skrajnie duża zmienność.

W związku z liczbą przedziałów zastosowano czterokrotnie zagnieżdżoną funkcję JEŻELI. Wynika to z faktu, że pojedyncza funkcja JEŻELI sprawdza wyłącznie jeden warunek, a zatem dwa przypadki – spełnionego i niespełnionego warunku. Formuła wyglądała następująco:

$$=JEŻELI(wsp. \text{zmienn.} < 0,2; "mała \text{ zmienność}";$$

$$JEŻELI(wsp. \text{zmienn.} < 0,4; "przeciętna \text{ zmienność}";$$

$$JEŻELI(wsp. \text{zmienn.} < 1; "duża \text{ zmienność}";$$

$$JEŻELI(wsp. \text{zmienn.} < 1,5; "bardzo duża \text{ zmienność}"; "skrajna \text{ zmienność}"))))$$

Rysunek 3.3 przedstawia stosowne wyliczenia w odniesieniu do przeprowadzonej analizy sprzedaży, z automatyczną interpretacją współczynnika zmienności dzięki zastosowaniu funkcji logicznej JEŻELI.

=JEŻELI(I18<0,2;"mała zmienność"; JEŻELI(I18<0,4;"przeciętna zmienność"; JEŻELI(I18<1;"duża zmienność"; JEŻELI(I18<										
	A	B	C	D	E	F	G	H	I	J
1						KP	KP			
2	dzień	mięsiąc rok	artykuł		jednostka	Kartuzy	Kwidzyn	Parametr	Kartuzy	Kwidzyn
3	30	Kwi	1996 Jabłka świeże	zł	23100,19126	28511,40414	średnia	15 251,82 zł	15 834,30 zł	
4	30	Cze	1996 Jabłka świeże	zł	9579,156216	14804,93428	min	8,16 zł	101,96 zł	
5	31	Lip	1996 Jabłka świeże	zł	11234,36205	7389,87732	max	29 968,88 zł	29 951,07 zł	
6	29	Lut	1996 Jabłka świeże	zł	25071,52852	18326,50241	wariancja	75046414	74154689	
7	31	Maj	1996 Jabłka świeże	zł	17096,26753	11871,47242	odchylenie standardowe	8 662,93 zł	8 611,31 zł	
8	31	Mar	1996 Jabłka świeże	zł	3792,604157	28337,78832	współczynnik zmienności	57%	54%	
9	31	Sie	1996 Jabłka świeże	zł	29748,41476	26106,76596	interpretacja	współczynnika zmienności	duża zmienność	duża zmienność

Rysunek 3.3. Analiza sprzedaży w oparciu o funkcje logiczne

Źródło: Opracowanie własne.

Zastosowanie instrukcji warunkowej, w tym przypadku funkcji JEŻELI, pozwoliło automatycznie zinterpretować współczynnik zmienności dla obu miast, jako „duża zmienność” (rysunek 3.3). Co istotne, interpretacja jest aktualizowana za każdym razem, gdy dokona się zmiana wartości w rekordach sprzedaży, dzięki czemu nie jest konieczne każdorazowe porównanie wartości współczynnika zmienności z wartościami referencyjnymi.

Wyniki analiz zrealizowanych w ramach niniejszego studium przypadku zawiera arkusz *analiza sprzedaży_2* w pliku *r3_analiza_sprzedaży_formuły.xlsx*.

3.3 Przypadek 3 – analiza danych struktury wykształcenia w oparciu o funkcje wyszukiwania i adresu

Firma X stale monitoruje rynek pracy oraz wykształcenie osób w różnych krajach pod kątem prowadzonych procesów rekrutacyjnych w poszczególnych regionach. W oparciu o uzyskane dane pochodzące z bazy danych Eurostatu i zaimportowane (rozdział 1 opracowania, przypadek 5) do arkusza MS Excel (plik *r1_wyższe_wykształcenie.xlsx*), przedsiębiorstwo pragnęło uzyskać informacje:

- w których krajach w poszczególnych latach było najwięcej osób z wyższym wykształceniem;
- dla każdego z krajów ustalić rok, w którym było najwięcej osób z najwyższym wykształceniem.

Powyższe informacje pozwoliłyby firmie nie tylko sprawdzić, który kraj ma najwięcej wykształconych osób, ale również dokonać analizy zachodzących zmian:

- wskazać kraje, które na przestrzeni lat były liderami w zakresie procentowej liczby osób z wyższym wykształceniem;

- określić kraje, w których najwyższa procentowa liczba osób z wyższym wykształceniem odnosi się do możliwie jak najbliższego okresu.

Wyszukiwanie jest integralną częścią przetwarzania danych wejściowych dla uzyskiwania użytecznych informacji [Deitel, Deitel, 2014]. W wielu przypadkach konieczne jest znalezienie powiązanych danych, np. wysokość podwyżki zgodnie z tabelą, zależnie od aktualnego wynagrodzenia całkowitego. Indywidualne, manualne wyszukiwanie powiązanych danych dla uzyskania informacji, jak w prezentowanym studium przypadku, wymagałoby sporych nakładów czasowych dla wykonania stosownych czynności, jak też obarczone byłoby znacznym ryzykiem popełnienia błędu. W związku z tym należało skorzystać z funkcji wyszukiwania i adresu automatyzujących, a przez to przyspieszających i usprawniających proces wyszukiwania informacji. Najważniejsze z nich zaprezentowano w tabeli 3.4.

Tabela 3.4. Podstawowe funkcje wyszukiwania i adresu

Funkcja	Znaczenie	Składnia
ILE.WIERSZY	Zwraca liczbę wierszy dla danego zakresu.	ILE.WIERSZY(tablica)
LICZBA.KOLUMN	Zwraca liczbę kolumn w odwołaniu.	LICZBA.KOLUMN(tablica)
WYSZUKAJ	Wyszukuje wartości w wektorze lub tablicy.	WYSZUKAJ(szukana_wartość; przeszukiwany_wektor; [wektor_wynikowy])
WYSZUKAJ.PIONOWO	Przeszukuje pierwszą kolumnę tablicy i przechodzi wzdłuż wiersza, aby zwrócić wartość stosownej komórki.	WYSZUKAJ.PIONOWO(szukana_wartość; tabela_tablica; nr_kolumny; [przeszukiwany_zakres])
WYSZUKAJ.POZIOMO	Przegląda górny wiersz tablicy i zwraca wartość wskazanej komórki.	WYSZUKAJ.POZIOMO(szukana_wartość; tabela_tablica; nr_indeksu_wiersza; [przeszukiwany_zakres])

Źródło: Opracowanie własne.

Na potrzeby analizy wynagrodzeń za użyteczną uznano funkcję WYSZUKAJ.PIONOWO, która pozwoliła wskazać kraje o najwyższym odsetku osób z wyższym wykształceniem, oraz WYSZUKAJ.POZIOMO, dla określenia lat, w których poszczególne kraje miały najwięcej osób z najwyższym wykształceniem. W przypadku pierwszej funkcji formuła miała następującą postać: =WYSZUKAJ.PIONOWO(MAX(zakres wartości dla konkretnego roku); zakres tabeli; numer kolumny, w której znajdują się nazwy krajów). W związku z tym, że szukana wartość (najwyższy odsetek osób z wyższym wykształceniem) jest zawsze w pierwszej kolumnie zakresu tabeli, nazwy krajów musiały zostać przeniesione na koniec zakresu tabeli. Wynik zastosowania formuły prezentuje rysunek 3.4.

K54															=WYSZUKAJ.PIONOWO(MAX(K13:K51);K13:\$L51;\$K12-K12+2;)		
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O		
32	Latvia	16,9	17,4	18,5	20,7	21,4	22,6	23,6	25,2	27,0	26,9	Latvia	20	2013	2005		
33	Lithuania	22,4	22,0	23,7	25,3	25,5	26,9	27,9	28,6	29,8	31,4	Lithuania	21	2014	2006		
34	Luxembourg	23,0	20,5	22,7	23,7	30,2	30,3	31,7	33,4	35,2	39,6	Luxembourg	22	2014	2006		
35	Hungary	14,5	15,0	15,4	16,4	16,9	17,1	18,0	19,0	19,5	20,2	Hungary	23	2014	2005		
36	Malta	10,3	11,1	11,7	12,1	12,8	14,2	15,1	16,4	17,2	18,0	Malta	24	2014	2005		
37	Netherlands	26,2	26,2	26,7	27,8	28,4	27,7	27,9	28,6	29,3	29,8	Netherlands	25	2014	2005		
38	Austria	15,0	14,7	14,6	15,0	16,0	16,2	16,3	16,9	17,7	27,4	Austria	26	2014	2007		
39	Poland	13,9	14,9	15,7	16,5	18,1	19,4	20,3	21,5	22,6	23,8	Poland	27	2014	2005		
40	Portugal	11,1	11,7	12,0	12,6	13,1	13,9	15,5	16,7	17,6	19,7	Portugal	28	2014	2005		
41	Romania	9,1	9,6	9,9	10,7	11,2	11,9	12,9	13,5	13,8	14,2	Romania	29	2014	2005		
42	Slovenia	16,7	17,8	18,5	19,0	19,6	20,2	21,6	23,0	24,4	25,1	Slovenia	30	2014	2005		
43	Slovakia	11,4	11,9	11,9	12,3	13,4	15,1	16,4	17,0	17,7	18,1	Slovakia	31	2014	2005		
44	Finland	28,6	29,0	30,0	30,2	30,9	31,6	32,5	32,8	33,6	34,7	Finland	32	2014	2005		
45	Sweden	25,9	25,9	26,4	26,9	27,6	28,2	29,1	30,1	31,4	32,8	Sweden	33	2014	2005		
46	United Kingdom	26,7	27,6	28,6	28,7	30,0	31,6	33,2	34,6	35,6	36,6	Kingdom	34	2014	2005		
47	Iceland	24,9	24,0	24,8	25,5	26,5	26,3	27,4	28,5	29,3	30,0	Iceland	35	2014	2006		
48	Norway	28,4	28,2	29,2	30,2	30,7	31,4	32,1	33,0	34,2	36,1	Norway	36	2014	2006		
49	Switzerland	24,2	25,2	26,5	28,5	29,6	30,0	30,0	31,2	33,2	34,4	Switzerland	37	2014	2005		
50	Former Yugoslav		10,6	11,7	11,0	12,5	13,3	15,0	15,7	15,1	15,2	Yugoslav	38	2012	2006		
51	Turkey		8,4	8,8	9,5	10,0	10,5	11,3	12,4	13,0	13,6	Turkey	39	2014	2006		
52	GEO/TIME	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	GEO/TIME		największy %	najmniejszy %		
53																	
54	Największy %	Finland	Denmark	Finland	Cyprus	Ireland	Ireland	Cyprus	Cyprus	Ireland	Luxembourg						
55	Najmniejszy %	Romania	Turkey	Turkey	Turkey	Turkey	Turkey	Turkey	Turkey	Turkey	Turkey						

Rysunek 3.4. Analiza odsetka osób z wyższym wykształceniem dla poszczególnych krajów

Źródło: Opracowanie własne.

Wyniki uzyskane dzięki funkcji WYSZUKAJ.PIONOWO wskazują, że kraj, w którym dla ostatniego monitorowanego roku (2014) było procentowo najwięcej wykształconych osób, to Luksemburg (rysunek 3.4). Natomiast kraje, które najczęściej były liderami w tym rankingu, to Cypr i Finlandia.

W przypadku drugiej stosowanej funkcji ustalenia dla każdego z krajów roku, w którym odsetek osób z najwyższym wykształceniem był największy, formuła miała postać: =WYSZUKAJ.POZIOMO(MAX(zakres wartości dla kraju); zakres tabeli; numer wiersza, w której znajdują się wartości lat). W związku z tym, że wyszukiwana była wartość w pierwszym wierszu podanego zakresu, lata zostały przeniesione na koniec zakresu. Wynik zastosowania formuły przedstawia rysunek 3.4. Zgodnie z nim w zdecydowanej większości krajów najwyższy procent liczby osób z wyższym wykształceniem występuje w roku 2014, czyli ostatnim badanym. Nielicznymi wyjątkami są takie kraje, jak Łotwa i FYROM (Former Yugoslav Republic of Macedonia).

Wyniki analiz zrealizowanych w ramach niniejszego studium przypadku zawiera plik *r3_wyzsze_wyksztalzenie_formuly.xlsx*.

3.4 Przypadek 4 – analiza danych sprzedaży z wykorzystaniem funkcji bazodanowych

W ramach pierwszego studium przypadku Firma X działająca w kilku miastach w Polsce dokonywała wyboru lokalizacji dla kolejnej hurtowni owoców. W celu

podjęcia decyzji wyliczane i interpretowane było m.in. odchylenie standardowe od wartości sprzedaży, które pozwoliło określić, jak duże było zróżnicowanie dla poszczególnych miesięcy. Dane sprzedaży w arkuszu (*dane sprzedaży*) skoroszytu (*r3_analiza_sprzedazy_formuly.xlsx*) nie były ułożone chronologicznie i przedstawiały sprzedaż wyłącznie dla pojedynczych miesięcy oraz większej liczby typów produktów niż analizowane. Nie były to wystarczające informacje przed podjęciem decyzji o wyborze miasta dla inwestycji. Firma chciała poznać dla badanych miast – Kartuz i Kwidzyna – jak wyglądała sprzedaż w okresie rocznym, uzyskując wskaźniki:

- średnia wartość sprzedaży,
- wartość sprzedaży za ostatni rok w podziale na jabłka świeże i gruszki świeże,
- maksymalna wartość sprzedaży w ostatnim roku jabłek świeżych i gruszek świeżych.

W prezentowanym studium przypadku dane musiały być analizowane w sposób przekrojowy [Drukker, 2011]. Wartość sprzedaży miała zostać obliczona dla wybranych lat oraz miast i towarów, przy równoczesnym przechowywaniu wartości obrotów na poziomie miesięcy. Stosowanie wyłącznie formuł z funkcjami matematycznymi i wyszukiwania było niewystarczające. Konieczne było przetworzenie danych z wykorzystaniem filtrowania dla wielu kolumn, w powiązaniu z funkcjami matematycznymi czy statystycznymi, w zintegrowany i kompleksowy sposób, typowy dla baz danych. W tym zakresie MS Excel posiada kilkanaście funkcji w kategorii bazy danych, z najważniejszymi zestawionymi w tabeli 3.5, z których wybrane znalazły zastosowanie dla wyliczenia pożądaných dwóch wskaźników.

Tabela 3.5. Wybrane funkcje bazodanowe

Funkcja	Znaczenie	Składnia
BD.ILE. REKORDÓW	Zwraca liczbę niepustych komórek znajdujących się w kolumnie lub bazie danych, które są zgodne z warunkami określonymi przez użytkownika.	BD.ILE.REKORDÓW(baza danych; pole; kryteria)
BD.MAX	Zwraca największą wartość z wybranych wpisów bazy danych.	BD.MAX(baza danych; pole; kryteria)
BD.MIN	Zwraca najmniejszą wartość z wybranych wpisów bazy danych.	BD.MIN(baza danych; pole; kryteria)
BD.SUMA	Zwraca sumę liczb w kolumnie lub bazie danych, które są zgodne z warunkami określonymi przez użytkownika.	BD.SUMA(baza danych; pole; kryteria)
BD.ŚREDNIA	Zwraca wartość średniej wybranych wpisów bazy danych.	BD.ŚREDNIA(baza danych; pole; kryteria)

Źródło: Opracowanie własne.

W związku z tym, że wyliczenia odnosiły się do wybranych rekordów w zależności od roku czy produktu, powiązано funkcje statystyczne z filtrowaniem. Do tego celu posłużyły funkcje bazodanowe:

- BD.ŚREDNIA(zakres danych tabeli; nazwa kolumny, czyli miasta; tabela z warunkami),
- BD.MAX(zakres danych tabeli; nazwa kolumny, czyli miasta; tabela z filtrem).

W tabeli z warunkami określono stosowne wartości filtra, czyli rok oraz nazwę produktu. Rezultaty zastosowania obu formuł zawiera rysunek 3.5.

G405		=BD.ŚREDNIA(\$C\$2:\$G\$400;"Kwidzyn";\$C412:\$C413)					
	A	B	C	D	E	F	G
394	31	Paź	2014	Jabłka świeże	zł	17904,64877	18632,04812
395	31	Sie	2014	Gruszki świeże	zł	1616,086765	25530,5914
396	31	Sie	2014	Jabłka świeże	zł	14355,77774	22249,20359
397	31	Sty	2014	Gruszki świeże	zł	20192,31946	18636,88187
398	31	Sty	2014	Jabłka świeże	zł	29959,91649	26893,08487
399	30	Wrz	2014	Gruszki świeże	zł	10008,72109	5050,455168
400	30	Wrz	2014	Jabłka świeże	zł	4525,713339	24295,4631
401							
402							
403				<i>Kartuzy</i>		<i>Kwidzyn</i>	
404				Jabłka świeże	Gruszki świeże	Jabłka świeże	Gruszki świeże
405			średnia	17 720,69 zł	12 065,40 zł	20 813,43 zł	14 761,16 zł
406			max	29 959,92 zł	24 593,43 zł	29 603,67 zł	25 530,59 zł
407							
408							
409	1	Jabłka	rok	<i>artykuł</i>			
410			2014	Jabłka świeże			
411							
412	2	Gruszki	rok	<i>artykuł</i>			
413			2014	Gruszki świeże			

Rysunek 3.5. Analiza sprzedaży w oparciu o funkcje bazy danych

Źródło: Opracowanie własne.

Otrzymane informacje przy zastosowaniu funkcji matematycznych, statystycznych i logicznych dane w ramach studiów przypadków pierwszego i drugiego wskazywały na bardzo niewielkie różnice w poziomie sprzedaży i jej zmienności dla poszczególnych miesięcy pomiędzy Kartuzami i Kwidzynem. Uzyskane dzięki funkcjom bazy danych wskaźniki, zaprezentowane na rysunku 3.5, poszerzają wcześniejsze informacje. Wskazują, że w ostatnim badanym roku sprzedaż jabłek świeżych była wyższa w Kwidzynie niż w Kartuzach (20 813 zł względem 17 721 zł).

Jeszcze większa różnica występuje w średniej wartości sprzedaży gruszek świeżych. Dla Kwidzyna była ona o 22% wyższa (14 761 zł względem 12 065 zł). Zatem ten rynek należy uznać za większy. W związku z uzyskanymi informacjami, kierując się wielkością rynku jabłek i gruszek dla ostatniego badanego roku, korzystniejsze jest zainwestowanie w Kwidzynie. Podjęcie decyzji wymaga

jednak analizy szerszego spektrum czynników, w tym takich, jak: nasycenie rynku czy warunki organizacyjno-finansowe wejścia na rynek.

Wyniki analiz zrealizowanych w ramach niniejszego studium przypadku zawiera arkusz *analiza sprzedaży_4* w pliku *r3_analiza_sprzedazy_formuly.xlsx*.

3.5 Przypadek 5 – analiza danych sprzedaży z wykorzystaniem funkcji finansowych

W ramach pierwszego, drugiego i czwartego studium przypadku Firma X działająca w kilku miastach w Polsce dokonywała wyboru lokalizacji dla nowego oddziału sprzedaży owoców – jabłek i gruszek. Do podjęcia decyzji wyliczany i interpretowany był m.in. współczynnik zmienności dla oceny zróżnicowania sprzedaży dla poszczególnych miesięcy (plik *r3_analiza_sprzedazy_formuly.xlsx*). Z punktu widzenia wyłącznie wielkości rynku i jego zmienności ustalono, że korzystniejsza inwestycja będzie w Kwidzynie. Kolejnym etapem przed podjęciem decyzji o rozpoczęciu inwestycji było określenie jej zyskowności. W tym celu przeprowadzona została stosowna analiza finansowa.

Analiza finansowa jest wraz z oceną techniczno-ekonomiczną kluczową składową analizy ekonomicznej stanowiącej filar badania działalności przedsiębiorstwa [Lee i in., 2009]. Celem analizy finansowej jest dostarczenie informacji o wynikach i sytuacji finansowej przedsiębiorstwa, niezbędnych w procesie zarządzania firmą. W zakres analizy finansowej wchodzi: weryfikacja sprawozdań finansowych czy ocena przepływów pieniężnych. Służą one ustaleniu sytuacji finansowej przedsiębiorstwa, w szczególności odnośnie do: płynności, zadłużenia, aktywności czy efektywności [Banerjee, 2010]. W tym zakresie MS Excel oferuje szerokie spektrum funkcji, z których najważniejsze zestawiono w tabeli 3.6.

Tabela 3.6. Wybrane funkcje finansowe

Funkcja	Znaczenie	Składnia
FV	Zwrócenie przyszłej wartości inwestycji.	FV(stopa; liczba_okresów; rata; [wb]; [typ])
PV	Obliczenie wartości bieżącej inwestycji.	PV(stopa; liczba_okresów; rata; [wp]; [typ])
NPV	Określenie bieżącej wartości netto inwestycji na podstawie szeregu okresowych przepływów gotówkowych oraz stopy dyskontowej.	NPV(stopa; wartość1; [wartość2]; ...)
RATE	Zwrócenie wysokości oprocentowania dla pojedynczego okresu raty rocznej.	RATE (liczba_okresów, rata, wb; [wp]; [typ], [wynik])

Funkcja	Znaczenie	Składnia
PMT	Obliczenie kwoty spłaty pożyczki przy założeniu stałych spłat i stopy procentowej.	PMT(stopa; liczba_okresów; wb; [wp]; [typ])
IRR	Zwrócenie wewnętrznej stopy zwrotu dla serii przepływów gotówkowych.	IRR(wartości; [wynik])

Źródło: Opracowanie własne.

Firma zakładała początkowe uzyskanie 25% udziału w rynku jabłek i gruszek. Zgodnie z wynikami czwartego studium przypadku (rysunek 3.5), oznaczało to miesięczny przychód na poziomie 8894 zł – ((średnia miesięczna sprzedaż jabłek + średnia miesięczna sprzedaż gruszek) * 25% udziału w rynku). Szacowana marża sprzedaży wynosiła 16,75% i jej połowa miała być reinwestowana w zwiększenie poziomu sprzedaży przez okres 2 lat. Miesięczne koszty stałe, jak wynajem budynku czy wynagrodzeń, wynosiły 8300 zł. Ponadto konieczne było zainwestowanie 11 000 zł w przygotowanie budynku. W związku z działalnością planowane było uzyskanie kredytu w wysokości 35 000 zł, który ma być spłacany przez okres 3 lat z comiesięczną ratą 1125 zł. Kredyt miał posłużyć adaptacji budynku, zakupowi urządzeń oraz zabezpieczeniu finansowania wynajmu lokalu i wynagrodzeń na okres 1 miesiąca:

- 7404 zł wydatków na produkty (jabłka i gruszki), jako 8894 zł (przychód ze sprzedaży) * 0,8325 (0,1675 stanowi marża),
- 8300 zł kosztów stałych,
- 11 000 zł na przygotowanie budynku.

Centrala firmy zagwarantowała środki na funkcjonowanie hurtowni przez okres 2 lat, czyli pokrywanie ewentualnych strat. Znając powyższe parametry, Firma X chciała określić, czy i w jakim stopniu prognozowane miesięczne przychody z inwestycji pokrywałyby planowane koszty całkowite w pierwszym miesiącu oraz po okresie 2 lat, gdy skończą się zabezpieczone środki finansowe. W tym celu określono:

- oprocentowanie kredytu,
- planowane przychody z inwestycji dla pierwszego i 24 miesiąca,
- stopień pokrycia kosztów przychodami z inwestycji.

Oprocentowanie kredytu wyliczono za pomocą funkcji RATE i wyniosło ono 9,73% w skali roku – $RATE(3*12;-1125;35000)*12$, gdzie:

- $3*12$ – liczba okresów, czyli miesiące płaćenia rat kredytu,
- 1125 – comiesięczna rata,
- 35 000 – kwota kredytu w złotych.

W skali miesiąca oprocentowanie kredytu wynosiło 0,8%. Planowany dochód ze sprzedaży miał wynosić 1490 zł, czyli $8894 \text{ zł} * 0,1675$. Zatem był on wystar-

czający do pokrycia kosztów kredytu. Niestety, firma miała ponieść 8300 zł kosztów stałych, co powodowałoby generowanie znacznych strat finansowych w pierwszym miesiącu, jak i wskazywało na nieopłacalność inwestycji.

Ambitne plany Firmy X zakładały, że połowa marży ze sprzedaży owoców będzie reinwestowana w zwiększenie obrotów. Dla określenia opłacalności prowadzenia biznesu po okresie 2 lat, gdy skończą się zabezpieczone środki na pokrycie strat, wyliczono poziom przychodów dla następných okresów, dla których prognozowano zwiększanie się sprzedaży. W tym celu wykorzystano funkcję finansową FV, która zwróciła prognozowaną wysokość przychodu jako 51 023 zł – $FV(0,1675/2;24;0;-7404)$, gdzie:

- 0,1675 – marża, o której połowę powiększany jest przychód dla kolejnych miesięcy dzięki zakupowi i sprzedaży większej liczby owoców;
- 24 – liczba okresów, czyli miesięcy, w trakcie których zwiększana jest sprzedaż;
- 0 – cała kwota jest inwestowana na początku bez comiesięcznych wpłat;
- 7404 zł – kwota zł, która jest inwestowana w pierwszym okresie na zakup produktów – $8894 \text{ zł planowanego przychodu} \cdot 0,8325$, ponieważ 0,1675 stanowi marża.

Prognozowany zysk ze sprzedaży owoców wyniósł 8546 zł, czyli $51\,023 \text{ zł (przychód)} \cdot 0,1675$ (marża). Całkowite koszty wyniosły natomiast 9425 zł, jako $8300 \text{ zł kosztów stałych} + 1125 \text{ zł raty kredytu}$. Oznacza to, że hurtownia również po okresie 2 lat działalności, przy ambitnych założeniach zwiększenia poziomu sprzedaży w tym okresie, nie będzie w stanie wygenerować przychodów pokrywających koszty. Zatem otworzenie hurtowni należy uznać za nieopłacalne we wskazanych warunkach prowadzenia działalności.

3.6. Możliwe zastosowania MS Excel w biznesie w formie zadań

W celu zobrazowania możliwości przetwarzania danych za pomocą formuł oraz funkcji poprzez narzędzie MS Excel można posłużyć się poniższymi zadaniami:

1. Firma X chce sprawdzić, jak wygląda zmienność dla 4 walut mających wymiar międzynarodowy: EUR, USD, GBP i CHF. Pozwoli to wybrać walutę o najmniejszych wahanach kursowych. W tym celu, korzystając z arkusza z kursami walut (*r1_kursy_walut.xlsx*), należy wyliczyć dla nich średnią oraz odchylenie standardowe.

2. Firma X, która monitoruje zmienność kursu dla 4 walut mających wymiar międzynarodowy, chce w pełni zautomatyzować ich analizę. W tym celu, korzystając z arkusza z kursami walut (*r1_kursy_walut.xlsx*) i wyliczonymi zgodnie z punktem pierwszym średnią i odchyleniem standardowym, należy zdefiniować

stosowną funkcję logiczną. Będzie ona informowała o poziomie zmienności kursu waluty poprzez komentarze tekstowe.

3. Wykorzystując dane sprzedaży jabłek świeżych i gruszek świeżych w poszczególnych miastach w Polsce, Firma X chce dokonać analizy porównawczej rynków dla kolejnych dwóch miast – Koszalina i Kołobrzegu – celem wstępnej oceny opłacalności otwarcia nowej hurtowni (wykorzystując plik *r2_dane_sprzedazy_o.xlsx*). W tym celu należy zastosować stosowne funkcje: matematyczne, statystyczne, logiczne i bazodanowe.

4. Firma X w związku z utratą zagranicznego kontrahenta poszukuje nowych rynków zbytu. W tym celu chce znaleźć 3 kraje, do których sprzedaż produktu o nazwie „Jabłka świeże” jest największa. Korzystając z arkusza z danymi eksportu w 2014 roku (*r2_import_i_eksport_o.xlsx*) oraz stosowanych funkcji matematycznych i wyszukiwania, należy określić, jakie to kraje.

5. Firma konsultingowa z Japonii planuje otworzyć nowy oddział w Europie. Kluczowym wyznacznikiem dla wyboru kraju jest poziom wykształcenia jego obywateli, który determinuje rynek potencjalnych pracowników. W oparciu o dane zawarte w skoroszycie prezentującym procent osób z wyższym wykształceniem w Europie (*r1_wyzsze_wyksztalcenie.xlsx*) należy ocenić ogólny poziom wykształcenia. W tym celu, dla ostatniego roku, należy określić, ile krajów miało odsetek w przedziale:

- poniżej 10%,
- 10–20%,
- 21–30%,
- powyżej 30%.

Należy wykorzystać odpowiednie funkcje bazodanowe.

6. Osoba X, zarabiająca średnią krajową, planuje comiesięcznie oszczędzać 10% swojej płacy netto, na koncie oprocentowanym na 3,5% w skali roku, przez okres 12 lat. Chce dowiedzieć się, czy na koniec okresu inwestycji uzbiera 100 000 zł. Jeśli nie, to chciałaby określić, ile musi oszczędzać miesięcznie. Do tego celu powinny zostać wykorzystane odpowiednie funkcje finansowe.

Zakończenie

Niniejszy rozdział wykazał znaczne możliwości, jakie posiada narzędzie MS Excel w zakresie przetwarzania danych z wykorzystaniem funkcji. Wskazane i zintegrowane w poszczególnych studiach przypadków funkcje: matematyczne, statystyczne, logiczne, wyszukiwania i adresu, baz danych oraz finansowe pozwalają uzyskać syntetyczne i przekrojowe dane. Ich umiejętne połączenie pozwala uzyskać ważne informacje dla prowadzonej działalności biznesowej, takie jak:

ocena wartości rynku; wartość sprzedaży w podziale na okresy, produkty i lokalizacje; opłacalność planowanej inwestycji w zakładanych warunkach jej prowadzenia.

Przedstawione studia przypadków i prezentowane w ich ramach rezultaty jednoznacznie wskazują, że narzędzie MS Excel posiada funkcjonalność pozwalającą przeprowadzić zaawansowaną analizę danych. W oparciu o szereg funkcji, ujętych w kilkunastu kategoriach, możliwe jest zdefiniowanie złożonych formuł pozwalających na przeprowadzenie procesu wnioskowania. W przypadku prezentowanych studiów przypadku odnosiło się to przede wszystkim do porównania lokalnych rynków zbytu dla owoców. Ważne jest postrzeganie przeprowadzenia analizy danych poprzez formuły i funkcje jako całościowego procesu. Dlatego też w ramach studiów przypadku zastosowano kilkanaście funkcji z obszarów: matematyczne, statystyczne, bazy danych, logiczne i finansowe dla kompleksowego i zautomatyzowanego wnioskowania w zakresie wyboru rynku zbytu, jak i potencjalnej opłacalności inwestycji wejścia na niego.

Przed zastosowaniem funkcji dostępnych w MS Excel należy uprzednio zadbać o pozyskanie właściwych danych, co umożliwi, jak to zaprezentowano w rozdziale 1, narzędzie MS Excel wspierające import danych z szeregu różnych źródeł. Ponadto ważne jest zapewnienie odpowiedniej jakości danych i formy ich reprezentacji, czemu poświęcony jest rozdział 2 niniejszej pracy.

Bibliografia

1. Abramowitz M., Stegun I.A. (1964), *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*, Courier Corporation.
2. Banerjee B.K. (2010), *Financial Accounting: Concepts, Analyses, Methods And Uses*, PHI Learning Pvt. Ltd.
3. Cieciora M., Zacharski J. (2011), *Podstawy Probabilistyki z Przykładami Zastosowań w Informatyce*, Warszawa.
4. Coontz S.H. (2001), *Population Theories and their Economic Interpretation*, Routledge.
5. Deitel H.M., Deitel B. (2014), *Computers and Data Processing*, Academic Press.
6. Drukker D.M. (2011), *Missing Data Methods: Cross-Sectional Methods and Applications*, Emerald Group Publishing.
7. Gamut L.T.F. (1991), *Logic, Language, and Meaning: Intensional logic and logical grammar*, University of Chicago Press.
8. Jackson M., Staunton M. (2006), *Advanced Modelling in Finance using Excel and VBA*, John Wiley & Sons.
9. Lee A.C., Lee J.C., Lee C.F. (2009), *Financial Analysis, Planning & Forecasting: Theory and Application*, World Scientific.
10. McFedries J. (2013), *Excel Data Analysis: Your visual blueprint for analyzing data, charts, and Pivot Tables*, John Wiley & Sons.

11. McFedries P. (2013), *Excel 2013 Formulas and Functions*, Que Publishing.
12. Penza P.K., Bansa V.P. (2001), *Measuring Market Risk with Value at Risk*, John Wiley & Sons.
13. Rao B.M. (2009), *Some Properties of the Coefficient of Variation and F Statistics with Respect to Transformations of the Form X_k* , University of Cornell.

Rozdział 4

Graficzna prezentacja danych biznesowych

Bartosz Marcinkowski

Wstęp

Przetworzone i odpowiednio sformatowane dane pierwotne – wraz z towarzyszącą im zawartością wtórną zapewnioną przez funkcje – są przedmiotem wielokryterialnej analizy i prezentacji biznesowej. Celem niniejszego rozdziału jest omówienie kluczowych form i mechanizmów wizualizacji danych, ukierunkowanych na zwiększenie efektywności analizy.

Za podstawową formę publikowania danych o charakterze statystycznym uznawane są tabele [Pieniążek i in., 2014]. Spowodowane jest to w szczególności ich precyzją i niekwestionowanym potencjałem w zakresie szczegółowej analizy danych. Tabele najczęściej stanowią optymalny wybór, gdy prezentacja danych [Lozovsky, 2008]:

- ma na celu weryfikację bądź porównanie konkretnych wartości,
- wymaga dużej dokładności,
- opiera się na wartościach wyrażonych w różnych jednostkach miar.

Jednocześnie mechanizmy formatowania danych – w tym formatowanie warunkowe – umożliwiają zwiększenie atrakcyjności prezentacji danych, ułatwiają analizę ich zróżnicowania i wyróżnianie zjawisk o charakterze systematycznym.

Wraz ze wzrostem szczebla zarządzania w organizacji gospodarczej znaczenia nabiera agregacja i graficzna prezentacja danych w postaci wskaźników, wykresów czy map. Środki graficznej prezentacji danych łączą techniczne umiejętności obróbki, analizy i interpretacji nierzadko olbrzymich wolumenów nieuporządkowanych danych z zasadami estetyki, innowacyjnym designem i interakcją [Paradowski, 2011]. Są one preferowaną formą prezentacji danych, gdy [Lozovsky, 2008]:

- jest ona silnie ukierunkowana na przedstawianie faktów wynikających z ukształtowania danych,
- ma na celu ukazanie zależności pomiędzy licznymi wartościami.

Odpowiednio przygotowane wykresy charakteryzują się wysokimi walorami wizualnymi i znajdują zastosowanie w prezentacjach biznesowych, raportach czy sprawozdaniach. Pozwalają sprawnie analizować trendy badanych zjawisk, identyfikować czynniki o charakterze sezonowym oraz porównywać kilka zmiennych bez potrzeby przeprowadzania drobiazgowej analizy. Środki graficznej prezen-

tacji danych zyskują coraz większe znaczenie w świecie biznesowym. Przykładami takich tendencji może być ewolucja narzędzia Microsoft Project – w coraz większym stopniu opierającego funkcjonalność sprawozdawczą związaną ze wspieranymi projektami na raportach graficznych – czy też rozwój kokpitów menedżerskich. Te ostatnie można postrzegać w kategoriach cennego narzędzia raportującego dla menedżerów przedsiębiorstw z sektora MSP – nie mają oni bowiem dostępu do tak szerokiego spektrum informacji zarządczych jak korporacje, ani nie dysponują zespołami analityków w celu zbierania danych i ich interpretowania [por. Dudycz i in., 2014].

Należy przy tym zaznaczyć, że tendencja do wzbogacania przekazu elementami graficznymi w praktyce może okazać się szkodliwa, gdyż niewłaściwy dobór środka przekazu lub nadmierne wykorzystanie grafiki tylko utrudnia jego interpretację [por. Błaszowski, Szanser, 2011]. Nawet poprawnie zaprojektowane, interaktywne środki prezentacji nie spełnią swojego zadania, o ile nie są wzbogacone komponentem opisowym – wyjaśniającym i interpretującym zjawiska i wyniki ujęte na wykresach [Pirrello, 2010].

4.1. Przypadek 1 – wykorzystanie formatowania warunkowego na potrzeby analizy wynagrodzeń

Formatowanie warunkowe to mechanizm automatycznego przypisywania formatów graficznych komórkom arkusza kalkulacyjnego na bazie wartości ujętych w poszczególnych komórkach. Mechanizm ten jest w praktyce prosty do zastosowania na dużą skalę z uwagi na możliwość tworzenia nowych reguł dla zakresów komórek oraz powielania formatowania warunkowego z wykorzystaniem malarza formatów. Dostęp do popularnych wariantów formatowania, pogrupowanych w menu *Formatowanie warunkowe*, możliwy jest bezpośrednio ze wstążki *Narzędzia główne*. Wykorzystanie jednej z propozycji narzędzia jest o tyle wygodne, że nie wiąże się z koniecznością manualnego projektowania stylu wyróżniania komórek. Pełną kontrolę nad tworzeniem zaawansowanych reguł (tabela 4.1) umożliwia z kolei opcja *Nowa reguła*. Dowolnej komórce arkusza można przypisać jednocześnie kilka reguł formatowania warunkowego. W takiej sytuacji domyślnie stosowane są one niezależnie, a komórkom przypisywane są wszystkie formaty wynikające z poszczególnych reguł. Ewentualne konflikty reguł są rozstrzygane zgodnie z regułą pierwszeństwa – reguła występująca jako wcześniejsza na kompletnej liście reguł jest nadrzędna w stosunku do wszystkich sprzecznych reguł występujących później [Dodge, 2013].

Tabela 4.1. Typy reguł w formatowaniu warunkowym

Typ reguły	Charakterystyka	Podstawowa parametryzacja
Formatuj wszystkie komórki na podstawie ich wartości	Wspomaga analizę porównawczą wartości w poszczególnych komórkach arkusza kalkulacyjnego. Reguła szczególnie przydatna w prezentacjach biznesowych.	<ul style="list-style-type: none"> – styl formatowania (skala 2-kolorowa, skala 3-kolorowa, pasek danych, zestawy ikon) – zakresy przedziałów – format wypełnień/pasków/styl ikon
Formatuj tylko komórki zawierające	Najpowszechniej stosowany typ reguły o wysokim stopniu uniwersalności. Przydatna zarówno w wyróżnianiu komórek na bazie indywidualnych kryteriów użytkownika, jak i w sprawdzaniu poprawności.	<ul style="list-style-type: none"> – warunek (wartość komórki, z określonym tekstem, z datami występującymi, puste, niepuste, zawierające błędy, niezawierające błędów) – parametr (dotyczy 3 pierwszych warunków) – format
Formatuj tylko wartości sklasyfikowane jako pierwsze i ostatnie	Umożliwia sprawną identyfikację skrajnych wartości w zbiorowościach, zarówno w oparciu o kryteria nominalne, jak i względne.	<ul style="list-style-type: none"> – próg ilościowy lub procentowy – format
Formatuj tylko wartości powyżej lub poniżej średniej	Wspomaga zaawansowaną analizę zróżnicowaną zbiorowości, przyjmując jako punkt odniesienia średnią arytmetyczną wszystkich wartości objętych formatowaniem warunkowym.	<ul style="list-style-type: none"> – reguła odniesienia do średniej (powyżej, poniżej, równe lub powyżej, równe lub poniżej oraz n wartości odchylenia standardowego powyżej/poniżej dla $n \leq 3$) – format
Formatuj tylko wartości unikatowe lub zduplikowane	Ułatwia identyfikację wierszy zawierających powtarzające się wartości (w szczególności – wskutek błędu na etapie importowania danych) w rozbudowanych arkuszach.	<ul style="list-style-type: none"> – cecha wartości (unikatowe, zduplikowane) – format
Użyj formuły do określenia komórek, które należy sformatować	Umożliwia dostosowanie formatowania warunkowego do indywidualnych potrzeb użytkownika poprzez wyróżnianie komórek na bazie autorskiej formuły lub grupy formuł cząstkowych połączonych operatorami LUB bądź ORAZ.	<ul style="list-style-type: none"> – specyfikacja formuły – format

Źródło: Opracowanie własne.

Kontrolę licznych reguł w poszczególnych obszarach arkusza kalkulacyjnego bądź całym arkuszu zapewnia opcja *Zarządzanie regułami*, uwzględniona w menu formatowania warunkowego. W szczególności umożliwia ona ustalanie kolejności wywoływania reguł i opcjonalnego zatrzymania walidacji kolejnych reguł na dowolnym etapie. Analogicznie, formatowanie warunkowe może koegzystować ze stylami manualnie narzuconymi przez użytkownika arkusza – w takiej sytuacji formatowanie warunkowe ma priorytet, a formaty manualne są przypisywane komórkom niespełniającym żadnej z reguł formatowania warunkowego.

Reguły formatowania warunkowego wykorzystano dla potrzeb analizy przeciętnych wynagrodzeń w wybranych krajach Europy w latach 2001–2008 (dane do analizy pozyskane zostały w rozdziale 1, przypadek 5 i skopiowane do pliku *r4_wynagrodzenia_brutto.xlsx*, arkusz *wynagrodzenia*) jak zaprezentowano na rysunku 4.1. Podstawową potrzebą analityczną w tym przypadku było wyszczególnienie wynagrodzeń w sposób istotny odbiegających od średniej badanej zbiorowości. Przyjęto skalę pięciostopniową, zgodnie z którą wartości poniżej/powyżej określonego ułamka średniej arytmetycznej traktowane są odpowiednio jako:

- < 20% – bardzo niskie,
- < 40% – niskie,
- < 60% – umiarkowane,
- < 80% – wysokie,
- \geq 80% – bardzo wysokie.

Z uwagi na konieczność zachowania czytelności formatowania także na monochromatycznych wydrukach, zrezygnowano ze skali kolorów na rzecz zestawów ikon.

KRAJ	2001		2002		2003		2004		2005		2006		2007		2008		
	€	zm %	€	zm %	€	zm %	€	zm %	€	zm %	€	zm %	€	zm %	€	zm %	
Belgia	-	-	-	-	33 873,00	-	-	-	-	-	-	-	-	-	-	-	-
Bulgaria	1 517,30	8%	1 638,10	8%	1 737,80	6%	1 859,90	7%	2 056,40	11%	2 269,10	10%	2 699,70	19%	3 415,00	26%	-
Dania	41 694,80	2%	43 577,30	5%	44 692,00	3%	46 122,00	3%	47 529,30	3%	48 307,30	2%	50 578,10	5%	52 703,10	4%	-
Niemcy	35 400,00	2%	36 500,00	3%	37 300,00	2%	38 000,00	2%	38 600,00	2%	39 066,00	1%	40 100,00	3%	41 200,00	3%	-
Irlandia	-	-	-	-	-	-	-	-	40 944,00	-	-	-	40 932,00	-	-	-	-
Grecja	16 630,00	5%	17 509,80	5%	18 122,20	3%	-	-	-	-	-	-	-	-	-	-	-
Hiszpania	17 873,80	-	18 580,30	4%	19 367,70	4%	20 018,60	3%	20 593,90	3%	21 724,70	5%	22 176,50	2%	-	-	-
Cypr	18 361,40	4%	19 396,50	6%	20 032,50	3%	20 541,20	3%	21 842,10	6%	22 836,40	5%	23 662,90	4%	25 579,00	8%	-
Węgry	-	-	-	-	6 929,30	-	7 395,60	7%	8 255,50	12%	8 280,00	0%	8 338,80	13%	-	-	-
Malta	14 267,00	6%	14 442,30	1%	14 294,60	-1%	14 522,00	2%	15 011,00	3%	15 593,00	4%	16 027,00	3%	16 313,00	2%	-
Austria	33 013,90	-	34 262,00	4%	35 186,00	3%	35 315,00	0%	36 533,00	3%	37 372,00	2%	38 458,70	3%	39 761,00	3%	-
Polska	-	-	-	-	-	-	6 278,70	-	-	-	-	-	-	-	-	-	-
Portugalia	13 309,00	6%	13 336,40	0%	13 491,40	1%	13 786,50	2%	14 068,40	2%	16 221,60	15%	17 201,60	6%	17 753,00	3%	-
Rumunia	-	-	-	-	-	-	2 429,00	-	3 218,00	32%	3 910,60	22%	5 044,20	29%	5 760,30	14%	-
Słowacja	3 619,90	6%	4 263,90	18%	4 644,20	9%	5 388,50	16%	6 047,20	12%	6 706,90	11%	8 031,30	20%	-	-	-
Finlandia	27 897,20	5%	29 060,00	4%	30 054,00	3%	31 055,00	3%	32 377,00	4%	33 069,00	2%	34 738,00	5%	36 625,00	5%	-
Szwecja	30 271,10	-4%	30 957,50	2%	32 034,00	3%	32 344,20	1%	32 902,30	2%	33 915,30	3%	35 534,00	5%	35 442,10	0%	-
Zjednoczone Królestwo	-	-	38 702,00	-	37 309,80	-4%	39 768,10	7%	-	-	42 654,70	-	43 675,40	2%	-	-	-
Islandia	34 099,70	-9%	35 233,00	3%	36 914,70	5%	41 362,00	12%	56 768,50	37%	55 891,10	-2%	64 042,00	15%	39 977,80	-38%	-

Rysunek 4.1. Wynagrodzenia w wybranych krajach Europy względem średniej z uwzględnieniem dynamiki wzrostu (arkusz *wynagrodzenia_form warun*)

Źródło: Opracowanie własne.

Jednocześnie w nawiązaniu do nominalnych danych płacowych wyliczono roczną dynamikę zmian płac w poszczególnych krajach w ujęciu procentowym i zastosowano formatowanie warunkowe do wyróżnienia wszystkich przypadków, w których dany kraj nie odnotował wzrostu przeciętnej płacy. W tym przypadku zastosowano autorski format, przewidujący kontrastowe kolory i pogrubioną czcionkę. Należy zwrócić uwagę, że w przypadku braku dostępności danych formuła nie przewiduje wstawiania wartości 0% – co wiązałoby się z popełnieniem błędu formalnego i zarazem inicjowałoby wyróżnienie dla przedmiotowych komórek.

Jako że obie zastosowane reguły odnoszą się do różnych kolumn tabeli, w rozpatrywanym przypadku nie występuje techniczna możliwość konfliktu pomiędzy obiema wyspecyfikowanymi regułami warunkowymi. Wyniki formatowania zrealizowanego w ramach niniejszego studium przypadku zamieszczono w arkuszu *wynagrodzenia_form warun* (plik *r4_wynagrodzenia_brutto.xlsx*).

4.2. Przypadek 2 – wykresy przebiegu w czasie jako alternatywna forma analizy dynamiki zjawiska w czasie

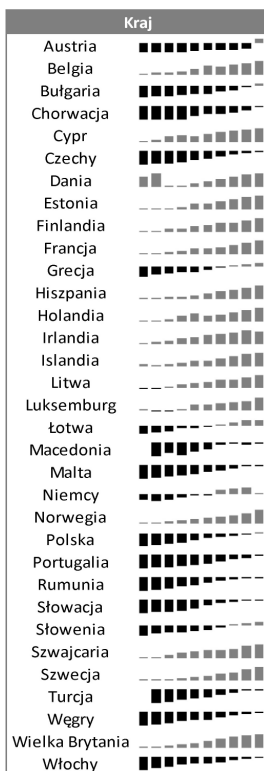
Wykresy przebiegu w czasie wzbogacają możliwości arkusza kalkulacyjnego w zakresie wizualizowania trendów i cykli ekonomicznych. Przyjmując formę zestawu zminiaturyzowanych wykresów ograniczonych rozmiarem do pojedynczych komórek arkusza kalkulacyjnego, uzupełniają funkcjonalność oferowaną przez formatowanie warunkowe. Podobnie jak formatowanie warunkowe można je sprawnie implementować na skalę masową, a zarazem dostosowywać do zindywidualizowanych potrzeb użytkownika zarówno układ graficzny, skalę, jak i opcje wykresów dla wielu różnych zestawów danych naraz [por. Duggirala, 2010]. Takiej elastyczności edycyjnej nie oferują tradycyjne wykresy. Wyróżnić można trzy typy wykresów przebiegu w czasie:

- liniowy,
- kolumnowy,
- zysk/strata.

Dobrą praktyką jest umieszczanie wykresów przebiegu w czasie bezpośrednio obok danych źródłowych, na podstawie których zostały one opracowane.

Na rysunku 4.2 zilustrowano trend w zakresie liczby osób z wyższym wykształceniem w różnych krajach europejskich na przestrzeni dekady (dane do analizy pozyskane zostały w rozdziale 1, przypadek 5 i skopiowane do pliku *r4_wyzsze_wyksztalzenie.xlsx*, arkusz *wyksztalzenie*). Wskaźniki w poszczególnych okresach przedstawiono jako odchylenia od wartości średniej (która w omawianym przypadku wynosi 22,72%) dzięki zastosowaniu typu wykresu przebiegu w czasie „zysk/strata”. Umożliwia to podkreślenie ujemnych odchyień od średniej ciemniejszą tonacją kolorystyczną, podczas gdy dodatnie odchylenia szkicowane są w notacji jaśniejszej. Z uwagi na cel graficznej prezentacji danych, tj. nakreślenie lokalnej dynamiki w poszczególnych krajach, zastosowano zindywidualizowaną skalę dla każdego wykresu składowego. Jeśli nadrzędną potrzebą byłoby przeprowadzenie analizy porównawczej siły zmian pomiędzy krajami, należałoby zastosować skalę ujednoliconą.

Wykresy przebiegu w czasie prezentujące trendy w wykształceniu Europejczyków w latach 2005–2014 zamieszczono w arkuszu *wyksztalzenie_wykr_przeb* (plik *r4_wyzsze_wyksztalzenie.xlsx*).



Rysunek 4.2. Trendy w wykształceniu Europejczyków w latach 2005–2014 (arkusz *wykształcenie_wykr_przeb*)

Źródło: Opracowanie własne.

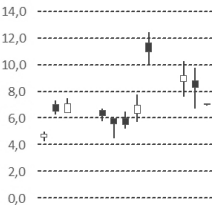
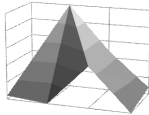
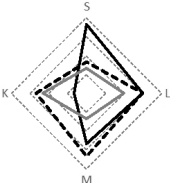
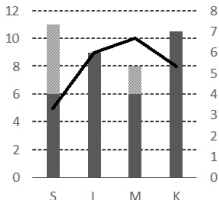
4.3. Przypadek 3 – graficzna prezentacja kursów walut w oparciu o różne odmiany wykresów tradycyjnych

Arkusze kalkulacyjne charakteryzujące się znacznym udziałem rynkowym cechują się jednocześnie szerokim wsparciem dla graficznej prezentacji danych, umożliwiając opracowywanie wykresów cieszących się uznaniem w naukach ilościowych i ekonomicznych. MS Excel 2016 dodatkowo rozszerza katalog oferowanych wykresów tradycyjnych (tabela 4.2) o nowe wykresy specjalistyczne, tj.:

- mapy w formie drzewa,
- wielopoziomowe wykresy pierścieniowe,
- histogramy oraz szczególny przypadek histogramów w postaci wykresów Pareto,
- wykresy typu „skrzynka i wąsy”,
- wykresy wodospadowe.

Tabela 4.2. Odmiany wykresów najczęściej stosowane do wizualizacji danych biznesowych

Rodzaj	Charakterystyka i zastosowanie	Przykład
Kolumnowy	Służy do prezentacji i przeprowadzania analizy porównawczej wartości kilku kategorii w danym momencie, jak i śledzenia zmian jednej lub kilku wartości kategorii w czasie. Miarą jest wysokość słupka lub jego części.	
Liniowy	Przedstawia szeregi liczbowe za pomocą linii w układzie współrzędnych prostokątnych. Sprawdza się przede wszystkim w prezentacji trendów w czasie.	
Kołowy	Służy do ilustrowania proporcji zjawiska. W wariancie podstawowym wartości kategorii są przedstawiane jako procenty kompletnego obszaru o kształcie koła.	
Słupkowy	Alternatywna postać wykresu kolumnowego, w której miarą jest długość poziomego słupka lub jego części. Pozwala na efektywne porównanie kilku kategorii o długich nazwach.	
Warstwowy	Ilustruje zmiany w czasie, podkreślając skumulowaną wartość trendu oraz umożliwiając analizę relacji wartości poszczególnych kategorii tworzących trend w dowolnym czasookresie.	
Punktowy	Przedstawia zależności zachodzące pomiędzy wartościami kategorii w postaci co najmniej dwóch zestawów punktów rozmieszczonych w układzie współrzędnych prostokątnych. Sprawdza się w analizie korelacji zestawów wartości.	

Rodzaj	Charakterystyka i zastosowanie	Przykład
Giełdowy	Dedykowany graficznej prezentacji wahań cen akcji lub zjawisk naukowych następujących w regularnych czasookresach.	
Powierzchniowy	Wspomaga analizę kombinacji między dwoma zestawami danych poprzez trójwymiarową wizualizację funkcji dwóch zmiennych.	
Radarowy	Umożliwia porównanie wartości liczbowych zaprezentowanych w układzie kołowych osi, rozprowadzanych promieniście od centralnego punktu.	
Kombi	Połączenie dwóch lub więcej rodzajów wykresów z uwzględnieniem ewentualnie osi pomocniczej.	

Źródło: Opracowanie własne na podstawie [Pieniżek i in., 2014; Excel Team, 2015].

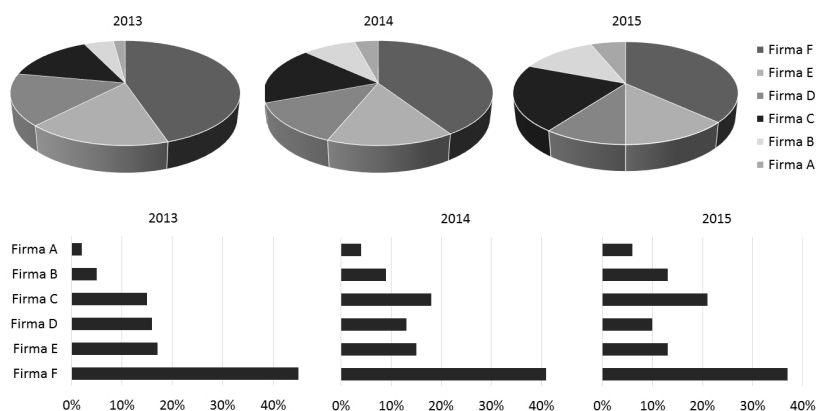
Wymienione nowo wprowadzone odmiany wykresów scharakteryzowano szerzej w tabeli 4.3 w ramach przypadku 4.

W praktyce najczęściej spotyka się wykresy kołowe, kolumnowe/słupkowe oraz liniowe. Szczególnie pierwszy z wymienionych jest często nadużywany w raportach biznesowych i materiałach marketingowych. T. Wolfram [2010] formułuje 10 argumentów za zaniechaniem powszechnego stosowania wykresów kołowych. Argumenty te obejmują m.in.:

- nieefektywność wykresów kołowych w zakresie porównywania danych;
- trudności w zakresie formatowania i późniejszego analizowania legendy oraz etykiet danych;
- niską skalowalność – najczytelniejsza postać wykresu kołowego ogranicza się do ilustrowania nie więcej niż 2–3 wartości;
- podatność wykresów kołowych na zaburzenia skali i błędy interpretacyjne;

- wynikające z historycznej popularności błędne przekonanie, że są one dobrym wyborem.

Do innych błędów często popełnianych podczas graficznej prezentacji danych ekonomicznych zaliczyć można skłonność do wykorzystywania trójwymiarowych wariantów wykresów (które koncentrują się na efektywności przekazu kosztem czytelności) oraz kontrastowej, bogatej kolorystyki uzupełnionej gradientami [por. Peltier, 2013]. Na rysunku 4.3 skonfrontowano nieefektywną formę prezentacji udziałów rynkowych grupy firm (odpowiednio w latach 2013, 2014 i 2015) w postaci trójwymiarowych wykresów kołowych z wizualizacją analogicznych danych na bazie uproszczonych wykresów słupkowych. Dla obu typów wykresów tradycyjnych zastosowano te same zbiory danych wejściowych w poszczególnych latach.



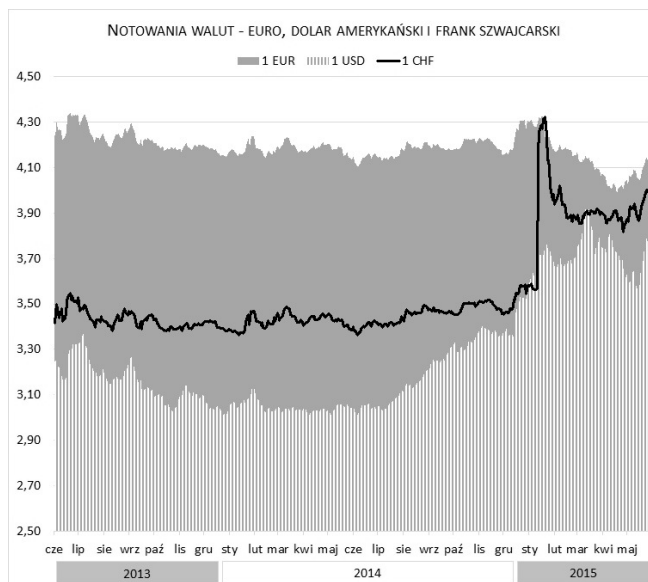
Rysunek 4.3. Przykład nieczytelnej i czytelnej formy śledzenia udziałów rynkowych

Źródło: Opracowanie własne na podstawie [Few, 2007].

Rysunek 4.4 poświęcono problematyce graficznej prezentacji kształtowania się kursów poszczególnych walut w czasie (dane pozyskane zostały w rozdziale 1, przypadek 3 i skopiowane do pliku *r4_kursy_walut.xlsx*, arkusz *kursy_wykresy*). Rysunek ten został oparty na wykresie kombi, integrującym różne rodzaje wykresów tradycyjnych z wykorzystaniem ujednoczonej osi wartości. Z uwagi na istotę omawianego przypadku oś pozioma (kategorii) ma charakter osi daty. Jako jednostkę główną przyjęto 1 miesiąc, aby zdefiniować czytelne jednostki osi o jednolitym odstępem czasowym. Zastosowano formatowanie niestandardowe daty – ograniczające się do wyłącznie do 3-literowego kodu miesiąca. Charakterystyka przypadku obejmuje bowiem bardzo liczne obserwacje. Aby uniknąć nagromadzenia nadmiarowych informacji na osi poziomej, na wykresie pogrupowano etykiety lat poprzez naniesienie cieniowanych pól tekstowych. W przypadku osi pionowej (wartości) zaakcentowano amplitudę wahań franka szwajcarskiego pomiędzy 14 a 15 stycznia 2015 roku poprzez ustalenie niestandardowego mini-

mum osi (2,5 zł). Jednostkę główną ustalono na 20 gr, podczas gdy jednostkę pomocniczą pominięto na wykresie.

Notowania franka szwajcarskiego przedstawiono w postaci wykresu liniowego, stanowiącego optymalny wybór w przypadku prezentowania śledzenia wartości zmiennych ciągłych bądź skokowych o znacznej liczbie obserwacji. Tło stanowią notowania EUR oraz USD. W obu przypadkach zastosowano wykresy warstwowe – odpowiednio o wypełnieniu jednolitym i deseniowym.



Rysunek 4.4. Kursy koszyka walut pomiędzy czerwcem 2013 r. a majem 2015 r. (arkusz *kursy_wykres*)

Źródło: Opracowanie własne.

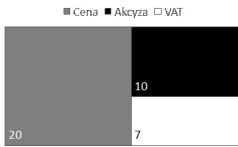
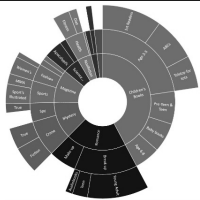
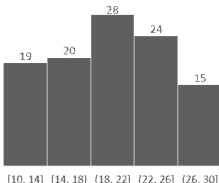
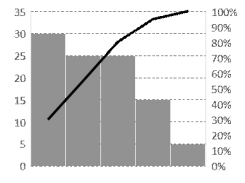
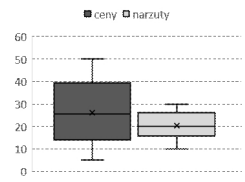
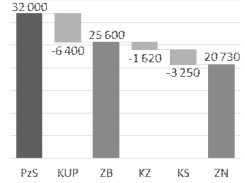
Wyniki graficznej analizy kursów koszyka walut w okresie czerwiec 2013–maj 2015 zamieszczono w arkuszu *kursy_wykres* (plik *r4_kursy_walut.xlsx*).

4.4. Przypadek 4 – wizualizacja sprzedaży z wykorzystaniem wykresów specjalistycznych

O ile wykresy specjalistyczne oferowane w arkuszu kalkulacyjnym MS Excel począwszy od wersji 2016 stanowią szczególną wartość dodaną przede wszystkim dla analityków biznesowych oraz praktyków nauk ilościowych, ich wprowadzenie w sposób zasadniczy poszerza także zestaw środków wyrazu, dostępnych twórcom kokpitów menedżerskich (tabela 4.3). Do tej pory w zakresie wizualizacji niestandardowej i zaawansowanej musieli oni bowiem polegać na dodatkach/rozwiąza-

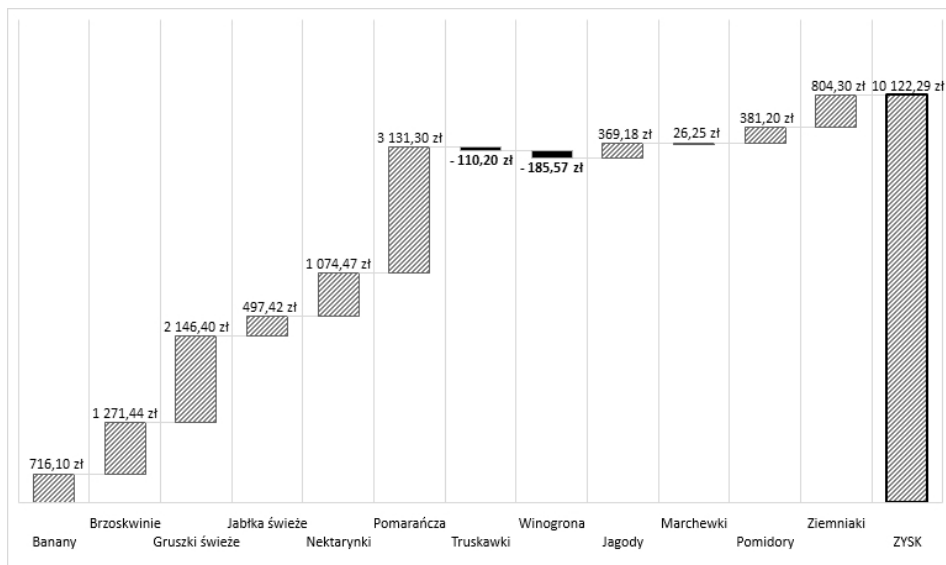
niach oferowanych przez strony trzecie, bądź przeznaczyć swój czas w dostosowanie dostępnej funkcjonalności prezentacyjnej do założeń projektowych kokpitu.

Tabela 4.3. Odmianny wykresów wprowadzone w arkuszu kalkulacyjnym MS Excel 2016

Rodzaj	Charakterystyka i zastosowanie	Przykład
Mapa w formie drzewa	Umożliwia porównywanie proporcji wartościowych w układzie hierarchicznym, podkreślając globalne udziały poszczególnych kategorii.	
Wielopoziomowy wykres pierścieniowy	Umożliwia porównywanie proporcji wartościowych w układzie hierarchicznym, podkreślając wzajemne relacje wartościowe pomiędzy hierarchicznie uporządkowanymi kategoriami i podkategoriami.	
Histogram	Przedstawia graficznie statystyczny rozkład zmiennej w układzie kolumnowym. Klasyfikacja i grupowanie wartości następuje automatycznie.	
Pareto	Szczególny przypadek uporządkowanego malejąco histogramu, podporządkowany regule 80/20. Podkreśla krytyczne czynniki mające wpływ na daną zmienną, wraz ze skumulowanym udziałem procentowym poszczególnych czynników.	
„Skrzynka i wąsy”	Umożliwia porównanie rozkładów statystycznych dwóch lub więcej zmiennych dzięki syntetycznemu ujęciu takich cech, jak mediana, kwartyły oraz wartość minimalna i maksymalna.	
Wodospadowy	Wizualizuje proces uzyskiwania wartości netto w drodze uznawania bądź obciążania wartości wyjściowej kwotami cząstkowymi. Użyteczny w tworzeniu sprawozdań finansowych.	

Źródło: Opracowanie własne na podstawie [Excel Team, 2015].

Dla potrzeb graficznej prezentacji i przeglądowej analizy składowych łącznego zysku/straty z tytułu sprzedaży produktów spożywczych w krytycznym czasookresie 2015 r. (dane źródłowe zawarte są w pliku *r4_faktury_sprzedazy_xlsx*, arkusz *zysk_strata*) posłużono się wykresem wodospadowym, który zaprezentowano na rysunku 4.5. Fakt wystąpienia w rozpatrywanym przypadku straty (która dotyczy dwóch grup asortymentowych) – a co za tym idzie, odnotowanie ujemnych wartości – utrudnia wizualizację w oparciu o tradycyjne wykresy skumulowane. Jak zaprezentowano na rysunku, konstrukcja wykresów wodospadowych umożliwia wyróżnienie wartości ujemnych oraz sum odmiennymi formatami. Z uwagi na domyślne stosowanie wartościowych etykiet danych na wykresach wodospadowych, wyeliminowano nadmiarową oś wartości oraz poziome linie siatki – z kolei zwiększając czytelność wykresu poprzez wprowadzenie pionowych linii siatki.



Rysunek 4.5. Kształtowanie się zysku/straty na sprzedaży produktów spożywczych względem grup asortymentowych (arkusz *zysk_strata_wykres*)

Źródło: Opracowanie własne.

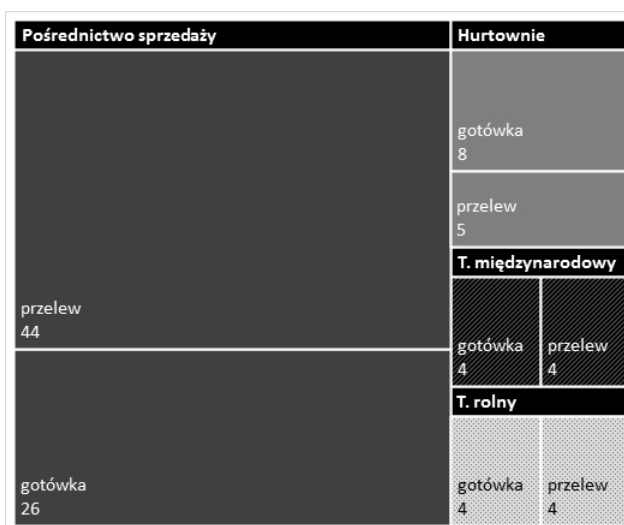
Wykres wodospadowy prezentujący kształtowanie się zysku/straty na sprzedaży produktów spożywczych zamieszczono w arkuszu *zysk_strata_wykres* (plik *4_faktury_sprzedazy_xlsx*).

Z kolei rozkład form płatności w poszczególnych kanałach sprzedaży (dane źródłowe zawarte są w pliku *r4_faktury_sprzedazy_xlsx*, arkusz *formy_płatności*) można syntetycznie podsumować z wykorzystaniem mapy w formie drzewa (rysunek 4.6). Zastosowanie dwóch serii danych sprawia, że racjonalne alternatywy stano-

wią w tym przypadku wykresy kolumnowe/słupkowe i wielopoziomowe wykresy pierścieniowe, również debiutujące w narzędziu MS Excel 2016.

Wykres dostosowano do indywidualnych potrzeb odbiorcy poprzez uzupełnienie etykiet danych o wartości umieszczone w osobnych wierszach bezpośrednio pod nazwami kategorii. Na potrzeby wydruku zrezygnowano również z tradycyjnej legendy wykresu na rzecz etykiet kategorii, wyświetlanych na mapie w formie drzewa w ujednoliconej kolorystyce.

Wykres typu mapa w formie drzewa prezentujący rozkład form płatności zamieszczono w arkuszu *formy płatności_wykres* (plik *r4_faktury_sprzedazy_xlsx*).



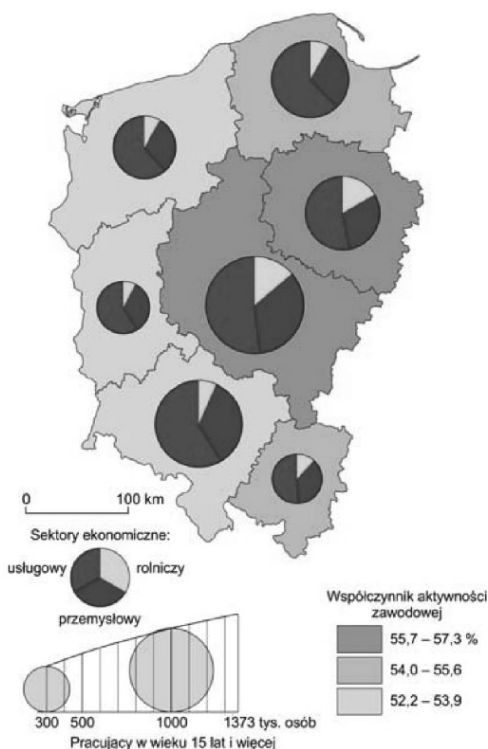
Rysunek 4.6. Rozkład form płatności w poszczególnych kanałach sprzedaży (arkusz *formy płatności_wykres*)

Źródło: Opracowanie własne.

4.5. Przypadek 5 – analiza porównawcza odsetka osób z wyższym wykształceniem z wykorzystaniem map

Szczególnym wyzwaniem z punktu widzenia graficznej prezentacji danych ekonomicznych jest wizualizacja danych w układzie terytorialnym. Dane takie są łatwiejsze do zinterpretowania dla potencjalnego odbiorcy, jeśli dysponuje on bezpośrednim nawiązaniem do podziału geograficznego, którego dotyczą prezentowane statystyki. Mapy tematyczne pozwalają na elastyczny dobór technik wizualizacji kilku zmiennych jednocześnie w oparciu o dane przestrzenne (por. rysunek 4.7).

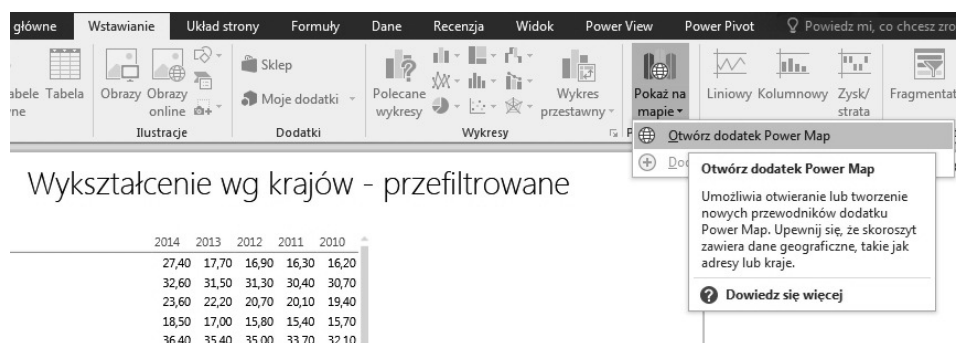
Mapy tematyczne znajdują zastosowanie przede wszystkim w graficznym prezentowaniu i analizie różnorodnych zjawisk – w tym geograficznych, ekonomicznych oraz demograficznych. W szerszym ujęciu mapy tematyczne sprawdzają się w ewidencji zasobów oraz infrastruktury. Wizualizacja danych przestrzennych z zastosowaniem map może stanowić podstawę do formułowania hipotez i generowania prognoz; kluczowe znaczenie ma w tym kontekście powiązanie wizualizacji z eksploracją danych, której istotą jest wydobycie wiedzy przestrzennej z wielkoskalowych zbiorów danych [por. Suhecka, 2014].



Rysunek 4.7. Prezentacja danych ekonomicznych z wykorzystaniem mapy tematycznej
 Źródło: [Pieniżek i in., 2014].

Mapy tematyczne w środowisku MS Excel 2016 bazują na mechanizmie raportowania Power View. Tym samym, po zapewnieniu zestawu danych wejściowych w postaci tabelarycznej (plik *r4_wyzsze_wyksztalzenie.xlsx*, arkusz *wyksztalzenie*), należy wstawić do arkusza raport dodatek Power View, korzystając ze wstążki *Wstawianie*. Należy w tym miejscu zaznaczyć, że mechanizm Power View wymaga, aby w systemie operacyjnym obecna była aktualna wersja Silverlight, czyli technologii prezentacji dedykowanej tworzeniu rozbudowanych interfejsów użytkownika, które mogą funkcjonować w dowolnym środowisku: w przeglądar-

kach internetowych, na różnych urządzeniach i w różnych systemach operacyjnych [szerzej: Moroney, 2007]. Przy pierwszym uruchomieniu Power View nastąpi ewentualny monit o zainstalowanie stosownej wersji technologii. Uporządkowane i przefiltrowane dane raportu są wizualizowane z wykorzystaniem dodatku Power Map (rysunek 4.8), który jest standardowo dostępny w arkuszu kalkulacyjnym Excel 2016.



Rysunek 4.8. Dodatek Power Map

Źródło: Opracowanie własne.

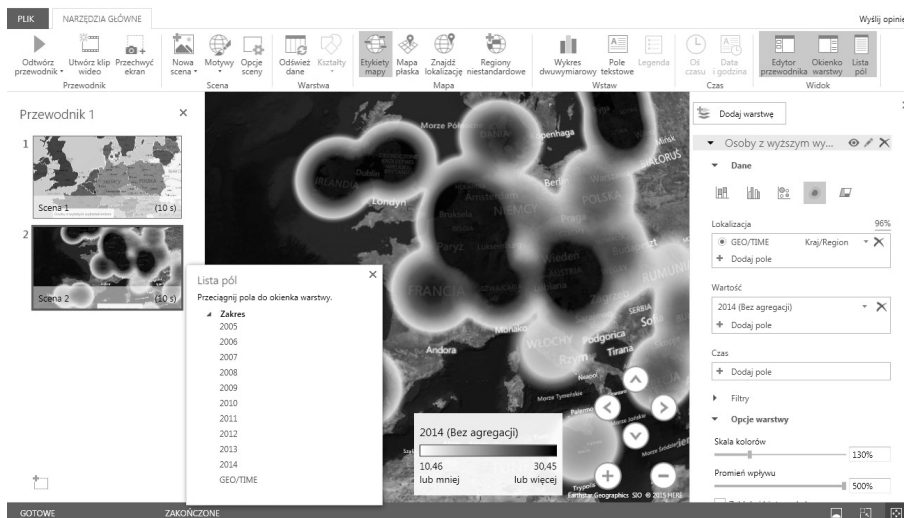
Realizacja niniejszego przypadku, tj. przeprowadzenie analizy porównawczej odsetka osób dysponujących wyższym wykształceniem z wykorzystaniem map, wymaga geokodowania. Proces ten przewiduje transformację danych wejściowych (tj. nazw krajów, regionów, miejscowości, kodów pocztowych itp.) na współrzędne geograficzne [por. Czapiewski, 2015] w oparciu o repozytoria danych przestrzennych i mapy Bing dostępne online. Skuteczność geokodowania na podstawie bazy danych Eurostatu wyniosła 96% – nie rozpoznano wpisu o składni *Former Yugoslav Republic of Macedonia, the*. Analityk może skorzystać z pięciu wariantów wizualizacji:

- skumulowanej kolumnowej – akcentującej składowe zjawiska,
- kolumnowej grupowanej – umożliwiającej porównywanie,
- bąbelkowej – zapewniającej dane porównawcze w oparciu o wykresy kołowe,
- konturowej – czyli tzw. „mapy ciepła”,
- regionalnej – tj. kartogramu, bazującego na tonacjach kolorystycznych.

Dwie pierwsze formy są ściśle ukierunkowane na przestrzeń trójwymiarową. Jak zaprezentowano na rysunku 4.9, na potrzeby późniejszej publikacji rozwiązania postawionego problemu zaprojektowano dwie sceny – pierwsza wykorzystuje wizualizację regionalną, podczas gdy druga – wizualizację konturową.

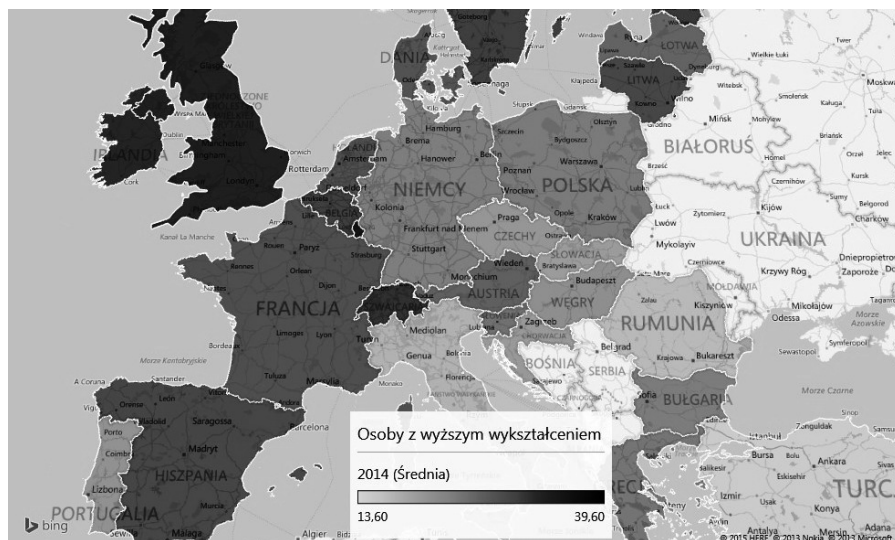
W obu scenach ograniczono się do pojedynczej warstwy danych. Wykorzystanie dodatkowych warstw pozwoliłoby analizować zjawisko względem kryteriów takich jak płeć lub przeprowadzać analizę porównawczą z innymi zmiennymi –

poziomem dochodów, preferencjami politycznymi itd. – co jest docelowym zadaniem sceny drugiej. Przeglądowa mapa tematyczna (rysunek 4.10) przyjmuje dla zwiększenia czytelności formę mapy płaskiej i wykorzystuje minimalistyczny motyw tła.



Rysunek 4.9. Power Map – parametryzacja scen

Źródło: Opracowanie własne.



Rysunek 4.10. Odsetek osób z wyższym wykształceniem w 2014 r. – rozkład geograficzny (arkusz *wykształcenie_mapa*)

Źródło: Opracowanie własne.

Mapę prezentującą odsetek osób z wyższym wykształceniem w 2014 r. w poszczególnych krajach Europy zamieszczono w arkuszu *wyksztalzenie_mapa* (plik *r4_wyzsze_wyksztalzenie.xlsx*).

4.6. Możliwe zastosowania MS Excel w biznesie w formie zadań

W celu zobrazowania możliwości wizualizacji danych z wykorzystaniem narzędzia MS Excel można posłużyć się poniższymi zadaniami:

1. Analityk danych odpowiedzialny za śledzenie sprzedaży produktów spożywczych planuje przeprowadzić analizę porównawczą wolumenu sprzedaży warzyw i owoców w poszczególnych tygodniach. Należy opracować taką analizę z wykorzystaniem zgrupowanych wykresów słupkowych.
2. Proszę opracować mapę ilustrującą w układzie dwuwymiarowym sumaryczną sprzedaż produktów spożywczych w poszczególnych województwach. Należy zastosować wykresy bąbelkowe, ujmujące odpowiednio: wolumen sprzedaży warzyw oraz wolumen sprzedaży owoców.
3. Bazując na danych wejściowych z przypadku 1, należy zastosować formatowanie warunkowe dla poziomów wynagrodzeń w postaci stopniowanych skali kolorów (3-stopniowych) oraz zaprezentować dynamikę zmian wynagrodzeń w poszczególnych krajach w postaci liniowych wykresów przebiegu w czasie.
4. Celem wzbogacenia prezentacji biznesowej o zaawansowaną animację należy zastosować dodatek Power Map do opracowania układu merytorycznie powiązanych scen w ramach pojedynczego przewodnika. Każda opracowana (skopiowana) scena winna odnosić się do kolejnego roku i przedstawiać graficznie poziomy wynagrodzeń w poszczególnych krajach europejskich w postaci wykresów kolumnowych w przestrzeni trójwymiarowej. Należy przetestować indywidualnie wybrane efekty przejścia w *Opcjach sceny* dla wybranej sceny i odtworzyć ją.
5. Firma X chce zbadać zmienność notowań funta szterlinga w porównaniu do euro, dolara amerykańskiego oraz franka szwajcarskiego, stosując podstawowe miary opisu statystycznego. Należy przedstawić wyniki stosownej analizy w oparciu o wykres typu „skrzynka i wąsy”.

Zakończenie

Niniejszy rozdział przedstawia analizę możliwości biznesowego zastosowania technik wizualizacji danych z wykorzystaniem arkusza kalkulacyjnego MS Excel.

Ich zakres pozwala na opracowywanie rozwiązań zarówno dla podstawowych, jak i zaawansowanych przypadków z zakresu nauk ekonomicznych i innych dyscyplin naukowych. Podkreślić należy stały wzrost potencjału narzędzia w tym zakresie, dzięki wprowadzaniu takich technik graficznej prezentacji danych, jak wykresy przebiegu w czasie, dodatkowe odmiany wykresów oraz funkcjonalność związana z prezentacją danych przestrzennych na mapach tematycznych.

Tematyka wizualizacji danych w narzędziu nie ogranicza się do aspektu statycznego – prezentując dane z wykorzystaniem dodatku Power Map, można tworzyć rozbudowane animacje, bazujące na szeregu wzajemnie powiązanych merytorycznie scen. Dzięki temu prezentacje biznesowe można wzbogacić o wielopoziomowe analizy porównawcze w układzie przestrzennym (w szczególności w układzie od-ogółu-do-szczegółu) w postaci plików wideo o wysokiej rozdzielczości. Przyczynia się to nie tylko do zwiększenia potencjału w zakresie analizy danych w organizacji gospodarczej, lecz także zwiększa atrakcyjność przekazu na spotkaniach biznesowych.

Bibliografia

1. Czapiewski B. (2015), *PowerMap, czyli potęga mapy (cz. 1) – Techniki przedstawiania danych*, www.powerview.pl.
2. Dodge M. (2013), *MOS 2013 Study Guide for Microsoft Excel Expert*, Microsoft Press, Redmond.
3. Dudycz H., Korczak J., Dyczkowski M. (2014), *Ontologiczna reprezentacja wiedzy finansowej w systemie wspomagania decyzji*, „Studia Ekonomiczne. Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach” nr 199.
4. Duggirala P. (2010), *What are Excel Sparklines & How to use them*, <http://chandoo.org/wp/2010/05/18/excel-sparklines-tutorial>.
5. Excel Team (2015), *Introducing new and modern chart types now available in Office 2016 Preview*, <https://blogs.office.com/2015/07/02/introducing-new-and-modern-chart-types-now-available-in-office-2016-preview>.
6. Few S. (2007), *Save the Pies for Dessert*, <http://www.perceptualedge.com/articles/08-21-07.pdf>.
7. Lozovsky V. (2008), *Table vs. Graph*, http://www.informationbuilders.com/new/newsletter/9-2/05_lozovsky.
8. Moroney L. (2007), *Wprowadzenie do Silverlight*, <https://msdn.microsoft.com/pl-pl/library/cc676565.aspx>.
9. Paradowski M. B. (2011), *Wizualizacja danych – dużo więcej, niż prezentacja*, w: Kluza M. (red.), *Wizualizacja wiedzy. Od Biblia Pauperum do hipertekstu*, Wiedza i Edukacja, Lublin.
10. Peltier J. (2013), *Excel Charting Dos and Don'ts*, <http://peltiertech.com/excel-charting-dos-and-donts>.

11. Pieniążek M., Szejgiec B., Zych M., Ajdyn A., Nowakowska G. (2013), *Graficzna prezentacja danych statystycznych. Wykresy, mapy, GIS*, Główny Urząd Statystyczny, Warszawa.
12. Pirrello C. (2010), *Effective Visualization Techniques for Data Discovery and Analysis*, w: *SAS Global Forum 2010 Proceedings*, SAS Institute, Cary.
13. Suhecka J. (red.) (2014), *Statystyka przestrzenna. Metody analizy struktur przestrzennych*, C.H. Beck, Monachium.
14. Wolfram T. (2010), *Countdown of Top 10 Reasons to Never Ever Use a Pie Chart*, https://blogs.oracle.com/experience/entry/countdown_of_top_10_reasons_to_never_ever_use_a_pie_chart.

Rozdział 5

Analiza danych z wykorzystaniem mechanizmów filtrowania i tabel przestawnych

Dorota Buchmowska

Wstęp

Skuteczne i efektywne zarządzanie organizacją gospodarczą wymaga podejmowania decyzji na podstawie danych (DDDM – ang. *Data Driven Decision Making*), a nie wyłącznie na podstawie intuicji [Provost, Fawcett, 2013]. Ze względu na ciągle rosnące zasoby informacji proces ten jest coraz trudniejszy i wymaga prowadzenia zaawansowanej analizy danych oraz zaangażowania przy tym technologii [Ziuziański, 2014]. Analityka biznesowa (określana również mianem *Business Intelligence* lub w szerszym ujęciu *Business Analytics*) rozumiana jako wydobywanie użytecznych informacji i wiedzy z dużych wolumenów danych w celu doskonalenia procesu podejmowania decyzji biznesowych [Provost, Fawcett, 2015] jest dla współczesnego przedsiębiorstwa determinantą wysokiej pozycji firmy na konkurencyjnym rynku.

Większość danych w przedsiębiorstwie pochodzi z wewnętrznych, transakcyjnych zbiorów danych, na których bazują systemy ewidencyjno-operacyjne, takie jak ERP (ang. *Enterprise Resource Planning*), CRM (ang. *Customer Relationship Management*) czy SCM (ang. *Supply Chain Management*). Systemy te automatyzują obsługę bieżącej działalności i pozwalają prowadzić ewidencję zdarzeń gospodarczych [Januszewski, 2008, s. 67]. Mimo iż rozwiązania te mają wbudowane podstawowe mechanizmy umożliwiające analizę danych – filtrowanie, tworzenie raportów, wykresów czy tabel przestawnych (ang. *pivot tables*) – większość użytkowników analizy danych dokonuje w arkuszu kalkulacyjnym, takim jak MS Excel. Zaletą tego narzędzia jest możliwość pobierania danych z różnych źródeł oraz ich integracji. Dane z baz transakcyjnych można importować do MS Excel za pośrednictwem mechanizmów opisanych w rozdziale 1. Natomiast możliwości integracji różnych zbiorów danych zaprezentowane zostały w rozdziale 6.

Dane pochodzące z baz transakcyjnych to jednak zazwyczaj dane szczegółowe, np. pojedyncze transakcje sprzedaży czy zakupu, gdy tymczasem menedżer potrzebuje danych zagregowanych, przekrojowych, takich jak całkowita wartość sprzedaży w danym okresie, całkowity zysk ze sprzedaży konkretnej grupy produktów. Dopiero wówczas można mówić o informacji zarządczej, wspierającej proces podejmowania decyzji [Surma, 2010, s. 49].

Przy analizie dużych zbiorów danych szczegółowych (w MS Excel nazywanych listami danych) sprawdzają się: sortowanie i sumy częściowe, filtrowanie danych oraz tabele i wykresy przestawne. Aby móc wykonywać operacje analityczne na liście danych, musi ona być odpowiednio przygotowana, tj.:

- ma posiadać wiersz nagłówkowy, czyli każda kolumna musi mieć nazwę (najlepiej unikatową), która wskazuje na jej zawartość;
- nie może zawierać pustych wierszy i kolumn;
- nie może zawierać scalonych komórek.

Celem niniejszego rozdziału jest zaprezentowanie rozwiązań umożliwiających szybką analizę dużych zbiorów szczegółowych danych z wykorzystaniem MS Excel 2016. W kolejnych studiach przypadków ukazane zostaną przykłady wykorzystania sortowania danych według różnych kryteriów i tworzenia sum częściowych, zaawansowanego filtrowania danych oraz analizowania danych za pośrednictwem interaktywnych tabel i wykresów przestawnych.

5.1. Przypadek 1 – wykorzystanie sum częściowych w analizie sprzedaży

Jednym z najczęstszych przedmiotów analizy danych jest sprzedaż towarów. Bez aktualnej wiedzy na temat przychodów i zysków generowanych przez określonych klientów lub sprzedaż poszczególnych grup towarów, monitorowania i ewaluacji pracy sprzedawców, czy też znajomości trendów i sezonowości dotyczących sprzedaży w zakresie asortymentu, firma nie jest w stanie się rozwijać. Dane dotyczące sprzedaży są łatwo dostępne dla każdej organizacji, ze względu na obowiązek prowadzenia ewidencji sprzedaży [PBSSP, 2015] oraz powszechność wykorzystywania systemów finansowo-księgowych, sprzedażowych, a także coraz częściej rozwiązań klasy ERP [GUS, 2014].

Firma X prowadząca sprzedaż hurtową owoców i warzyw, na podstawie wyników sprzedaży z ubiegłych okresów (arkusz *faktury sprzedaży*, plik *r5_faktury_sprzedazy_sumycz.xlsx*, zawierający skopiowane dane z pliku *r2_faktury_sprzedazy_no.xlsx*), co kwartał dokonuje oceny zyskowności oraz trendów w sprzedaży swoich produktów, by na tej podstawie prawidłowo zdefiniować strukturę asortymentową na najbliższy okres. Ze względu na rodzaj oferowanych towarów i związaną z nimi sezonowość sprzedaży analiza obejmuje nie tylko ostatnie miesiące, ale w zależności od sytuacji (między innymi od warunków pogodowych w danym roku) okres od jednego do dwóch lat. W lipcu 2015 roku dokonano oceny ilościowej i wartościowej sprzedaży za lata 2014–2015 w dwóch wymiarach: produkty oraz klienci. Do realizacji tego celu wykorzystane zostało na-

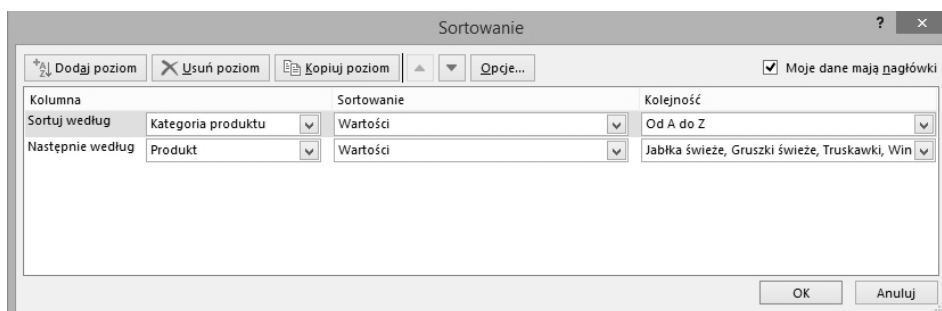
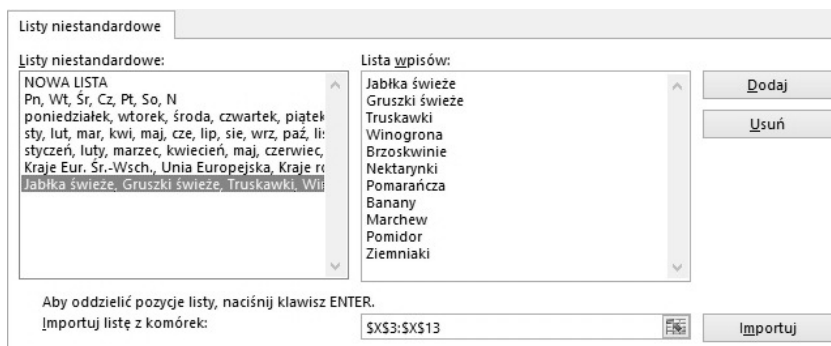
rzędzie MS Excel 2016. Analitycy biznesowi opracowali koncepcję systemu analitycznego, możliwego do realizacji w tym narzędziu.

Do wstępnej oceny sprzedaży pod kątem ustalenia konieczności wprowadzenia zmian w strukturze asortymentowej zastosowano narzędzie *Suma częściowa* (dostępne na wstążce *Dane*). Stosując je, obliczono sumaryczne przychody i zyski ze sprzedaży (wcześniej wartości te dla pojedynczych rekordów obliczono za pomocą stosownych formuł) poszczególnych produktów i grup produktów. Przed zastosowaniem *Sumy częściowej* konieczne było odpowiednie posortowanie danych (polecenie *Sortuj*, wstążka *Dane*), dlatego też w tym przypadku zdefiniowano dwa poziomy sortowania. Pierwszy – według kategorii produktów, żeby zsumować wartości przychodów i zysków dla poszczególnych kategorii (owoców i warzyw). Drugi – według poszczególnych produktów, żeby ustalić sumaryczne wielkości sprzedaży oraz przychody i zyski z konkretnych produktów (rysunek 5.1).

W przypadku drugiego klucza zastosowano niestandardową kolejność sortowania, dzięki czemu ułatwiona została analiza i ocena różnych grup towarów (w tym przypadku owoców krajowych i importowanych). Wymagało to utworzenia listy niestandardowej (*Opcje programu Excel*, kategoria *Zaawansowane*) o następujących kolejnych pozycjach: Jabłka świeże, Gruszki świeże, Truskawki, Winogrona, Brzoskwinie, Nektarynki, Pomarańcza, Banany, Marchew, Ziemniaki. Żeby uniknąć błędów literowych, listę z nazwami produktów stworzono poprzez skopiowanie ich nazw do osobnego zakresu, a następnie zaimportowanie tej listy (rysunek 5.1).

Po prawidłowym posortowaniu danych, za pośrednictwem narzędzia *Suma częściowa*, obliczone zostały sumy przychodów i zysków (ilości pominięto ze względu na różną jednostkę miary) dla kolumny *Kategoria produktu* (czyli głównego kryterium sortowania), a następnie w drugiej iteracji sumy z ilości, przychodów oraz zysków dla poszczególnych produktów. Efekt został zaprezentowany na rysunku 5.2. Należy pamiętać, że narzędzie *Suma częściowa* nie jest aktywne, jeżeli lista danych ma postać tabeli. Wówczas tabelę należy przekonwertować na zakres danych (wstążka *Projektowanie*, polecenie *Konwertuj na zakres*).

Jak widać, Firma X osiąga trzy razy większe przychody z owoców niż z warzyw, a różnica w zysku jest jeszcze wyższa. Owocami, które generują najwyższe przychody, są winogrona, jednak przynoszą one stosunkowo niskie zyski. Zdarza się, że przedsiębiorstwa podejmują decyzje biznesowe (np. dotyczące struktury asortymentowej) tylko na podstawie cząstkowych, najbardziej dostępnych danych (np. przychodów ze sprzedaży), gdy tymczasem zastosowanie nawet niezaawansowanych narzędzi pozwala na lepszą ocenę sytuacji i chroni przed podjęciem błędnych decyzji.



Rysunek 5.1. Sortowanie danych z wykorzystaniem list niestandardowych

Źródło: Opracowanie własne.

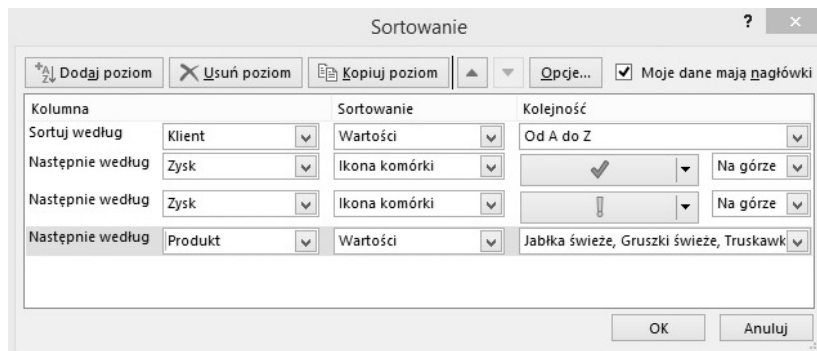
	B	C	D	E	F	G	H	I	L	M	N	S	T	
	1	Data	Produkt	Kategoria produktu	sprzedaży netto	Cena zakupu	Rabat	Ilość	Jednostka miary	Klient	Miejscowość	Województwo	Przychód	Zysk
	158		Jabłka świeże	Suma				17589					33 090,30 zł	26 054,70 zł
	410		Gruszki świeże	Suma				27072					102 362,80 zł	68 094,00 zł
	535		Truskawki	Suma				13134					146 519,73 zł	11 678,73 zł
	880		Winogrona	Suma				38343					293 964,48 zł	17 138,68 zł
	1225		Brzoskwinie	Suma				38907					91 591,58 zł	45 185,39 zł
	1569		Nektarynki	Suma				38469					109 212,45 zł	55 355,85 zł
	1913		Pomarańcza	Suma				38232					162 306,00 zł	82 385,00 zł
	2165		Banany	Suma				27719					89 104,14 zł	33 686,14 zł
	2186		Owoce	Suma									1 028 151,48 zł	339 538,48 zł
	2512		Marchew	Suma				38348					53 837,64 zł	14 846,24 zł
	2860		Pomidor	Suma				38149					180 547,50 zł	27 951,50 zł
	3112		Ziemniaki	Suma				26419					75 193,05 zł	22 355,05 zł
	3115		Marchewki	Suma				332					453,90 zł	85,90 zł
	3116		Warzywa	Suma									310 032,09 zł	65 238,69 zł
	3117		Suma końcowa					342713					1 338 183,57 zł	404 777,17 zł
	3118		Suma końcowa											

Rysunek 5.2. Efekt zastosowania narzędzia *Suma częściowa* (arkusz *sumy częściowe_1*)

Źródło: Opracowanie własne.

Drugim wymiarem, według którego dokonano wstępnej analizy danych, są klienci. Policzono sumaryczne i średnie wartości przychodów oraz zysków generowanych przez poszczególnych klientów. Im większa liczba i bardziej różnicowani klienci, tym więcej należałoby wziąć pod uwagę kryteriów ich grupowania (np. rodzaj klienta – biznesowy, indywidualny; położenie – województwo, miasto itd.), żeby zauważyć pewne zależności i tendencje. W przypadku Firmy X liczba klientów nie jest duża, są to klienci biznesowi i każdy z nich pochodzi z innego

miasta, analizie więc poddano każdego indywidualnie. Żeby łatwiej można było ocenić zyskowność poszczególnych produktów w przypadku każdego z klientów, dane posortowano dodatkowo według zysku ze sprzedaży oraz nazw produktów (rysunek 5.3).



Rysunek 5.3. Sortowanie danych według ikony komórki

Źródło: Opracowanie własne.

Sortowanie według zysku zostało przygotowane nie na szczegółowych wartościach, lecz trzech grupach wartości – w zależności od ich relacji do średniej. W tym celu zastosowano *Formatowanie warunkowe* (które omówiono w rozdziale 4), opcję *Zestawy ikon* i osobno wyróżniono wartości stanowiące $\geq 67\%$ średniej; $\geq 33\%$ średniej i poniżej 33% średniej. Następnie obliczono średni przychód i średni zysk dla poszczególnych klientów, wykorzystując narzędzie *Suma częściowa*. Efekt zaprezentowano na rysunku 5.4.

	B	C	D	E	G	H	L	M	N	S	T	
1	Data	Produkt	Kategoria	proc	Cena sprz	Rabat	Ilość	Klient	Miejscowość	Województwo	Przychód	Zysk
216	2014-11-10	Ziemniaki	Warzywa		3,00 zł	0%	56	ABC Sp. z o.o.	Wrocław	dolnośląskie	168,00 zł	56,00 zł
217	2015-01-19	Ziemniaki	Warzywa		3,00 zł	0%	33	ABC Sp. z o.o.	Wrocław	dolnośląskie	99,00 zł	33,00 zł
218	2015-03-12	Ziemniaki	Warzywa		3,00 zł	0%	35	ABC Sp. z o.o.	Wrocław	dolnośląskie	105,00 zł	35,00 zł
219	2015-03-31	Ziemniaki	Warzywa		3,00 zł	0%	55	ABC Sp. z o.o.	Wrocław	dolnośląskie	165,00 zł	55,00 zł
220	2015-04-17	Ziemniaki	Warzywa		3,00 zł	0%	59	ABC Sp. z o.o.	Wrocław	dolnośląskie	177,00 zł	59,00 zł
221	2015-05-22	Ziemniaki	Warzywa		3,00 zł	0%	41	ABC Sp. z o.o.	Wrocław	dolnośląskie	123,00 zł	41,00 zł
222								ABC Sp. z o.o. Średnia			327,09 zł	156,41 zł
380								Adam Nowak Sp.J. Średnia			468,66 zł	144,14 zł
537								Hurtownia dla vegetarian sp z o.o. Średnia			373,68 zł	88,26 zł
725								Hurtownia spozywca Średnia			402,17 zł	123,08 zł
977								Hurtownia spozywca w Białymstoku Średnia			662,59 zł	113,02 zł
1165								Jan Kowalski, Sp. J. Średnia			299,55 zł	104,53 zł
1355								Mazurskie wyroby Sp. z o.o. Średnia			406,93 zł	81,77 zł
1512								Multitybór owoców Sp. z o.o. Średnia			402,65 zł	147,18 zł
1702								Najlepsza żywność Sp. z o.o. Średnia			350,53 zł	100,65 zł
2019								Owoce i warzywa Sp. z o.o. Średnia			410,75 zł	129,39 zł
2208								Owoce z Zachodniego Pomorza, Sp. z o.o. Średnia			379,05 zł	152,24 zł
2365								Wegan Hurt Sp z o.o. Średnia			473,95 zł	126,05 zł
2521								Zakłady spozywcze SA Średnia			523,51 zł	159,93 zł
2678								Zdrowa żywność SA Średnia			330,43 zł	164,19 zł
2962								Żywność Gdańska Sp. z o.o. Średnia			513,88 zł	159,46 zł
3118								Żywność hurt i detal Sp. z o.o. Średnia			501,50 zł	134,84 zł
3119								Średnia całkowita			431,53 zł	130,53 zł

Rysunek 5.4. Fragment wyniku zastosowania *Sumy częściowej* do obliczenia średniej z przychodów i zysku dla poszczególnych klientów (arkusz *suma częściowe_2*)

Źródło: Opracowanie własne.

W kolejnych krokach obliczono sumę przychodu i zysku dla poszczególnych klientów, a następnie liczbę transakcji. Pozwoliło to ocenić, który z klientów generuje najwyższe, a który najniższe przychody i zyski oraz który z klientów dokonuje najczęściej zakupów. Z wykorzystaniem *Sumy częściowej* można dokonać dalszej analizy, obliczając podstawowe miary statystyczne (np. odchylenie standardowe i wariancję). Pozwoli to określić stałość zachowań zakupowych poszczególnych klientów, co może być podstawą do podjęcia decyzji o potrzebie przeprowadzenia dalszych analiz, np. przyczynowo-skutkowych, pozwalających zidentyfikować czynniki wpływające na zmiany zakupów.

Efekty zrealizowanych w ramach niniejszego studium przypadku analiz zawarte są w pliku *r5_faktury_sprzedazy_sumycz.xlsx*.

5.2. Przypadek 2 – zastosowanie mechanizmów filtrowania w celu dostosowania danych do potrzeb informacyjnych decydentów

Częstą sytuacją przy dużych zbiorach danych jest nadmiarowość. Przed przystąpieniem do dalszej analizy należy dane przefiltrować celem wybrania tylko tych spełniających określone kryteria, dopasowanych do potrzeb informacyjnych decydentów. Wówczas dane takie określa się mianem relewantnych [Abramowicz, 2003]. Filtrowanie może być też wykorzystane w celu określenia dalszych metod badawczych, kolejnych etapów analizy, do wstępnej oceny zjawiska czy też oceny i poprawy jakości danych.

W Excelu dostępne są dwa sposoby filtrowania (wstążka *Dane*) – autofiltr (polecenie *Filtruj*) i filtr zaawansowany (polecenie *Zaawansowane*). Autofiltr jest prostym, intuicyjnym narzędziem, które pozwala na wybieranie rekordów według listy wartości, według formatu lub według kryteriów. W jednej kolumnie można zastosować w danym momencie tylko jeden rodzaj autofiltera. W pierwszym przypadku wystarczy zaznaczyć wybrane wartości z listy występujących w danej kolumnie. W przypadku dłuższej listy można skorzystać z polecenia *Wyszukaj*, które pozwala na korzystanie ze znaków globalnych („?” – zastępujący dowolny pojedynczy znak lub „*” – zastępujący dowolną liczbę znaków).

Do filtrowania danych według wartości można wykorzystać fragmentatory (ang. *slicers*). W tym celu należy przekonwertować listę danych na tabelę (wstążka *Wstawianie*, polecenie *Wstaw tabelę*), a następnie wstawić stosowne fragmentatory (wstążka *Wstawianie* lub *Projektowanie*) dla kolumn, według których dane mają być filtrowane.

Firma X wykorzystuje fragmentatory do wyszukiwania transakcji sprzedaży (dane w arkuszu *faktury sprzedazy*, plik *r5_faktury_sprzedazy_filtr.xlsx*) realizowanych

za pośrednictwem wybranych sprzedawców w określonych województwach, co przedstawia rysunek 5.5. W obu fragmentatorach zastosowano wybór wielokrotny. Zaletą ich wykorzystania jest widoczność nie tylko zastosowanych kryteriów, ale również wartości, które znalazły się w wynikach wyszukiwania i tych, które nie znalazły się w wynikach (arkusz *fragmentatory*).

Produkt	Kategoria produktu	Ilość	Jednostka miary	Sprzedawca	Klient	Miejscowość	Województwo	Przychód ze sprzedaży
Ziemniaki	Warzywa	176 kg	Targ rolny	ABC Sp. z o.o.	Wrocław	dolnośląskie	528	
Ziemniaki	Warzywa	86 kg	Targ rolny	Zywność hurt i def.Lódź		łódzkie	258	
Ziemniaki	Warzywa	31 kg	Targ rolny	Owoce z Zachodni Szczecin		zachodniopomorskie	93	
Ziemniaki	Warzywa	42 kg	Targ rolny	Owoce z Zachodni Szczecin		zachodniopomorskie	126	
Ziemniaki	Warzywa	158 kg	Targ rolny	ABC Sp. z o.o.	Wrocław	dolnośląskie	474	
Ziemniaki	Warzywa	159 kg	Targ rolny	Zywność hurt i def.Lódź		łódzkie	477	
Ziemniaki	Warzywa	200 kg	Targ rolny	Owoce z Zachodni Szczecin		zachodniopomorskie	600	
Ziemniaki	Warzywa	96 kg	Targ rolny	Owoce z Zachodni Szczecin		zachodniopomorskie	288	
Ziemniaki	Warzywa	67 kg	Targ rolny	ABC Sp. z o.o.	Wrocław	dolnośląskie	201	
Ziemniaki	Warzywa	46 kg	Targ rolny	Zywność hurt i def.Lódź		łódzkie	138	
Ziemniaki	Warzywa	129 kg	Targ rolny	Owoce z Zachodni Szczecin		zachodniopomorskie	387	
Ziemniaki	Warzywa	80 kg	Targ rolny	Owoce z Zachodni Szczecin		zachodniopomorskie	240	
Ziemniaki	Warzywa	105 kg	Targ rolny	ABC Sp. z o.o.	Wrocław	dolnośląskie	315	
Ziemniaki	Warzywa	153 kg	Targ rolny	Zywność hurt i def.Lódź		łódzkie	459	
Ziemniaki	Warzywa	41 kg	Targ rolny	Owoce z Zachodni Szczecin		zachodniopomorskie	123	
Ziemniaki	Warzywa	113 kg	Targ rolny	Owoce z Zachodni Szczecin		zachodniopomorskie	339	
Ziemniaki	Warzywa	132 kg	Targ rolny	ABC Sp. z o.o.	Wrocław	dolnośląskie	396	

Województwo		
dolnośląskie	lubuskie	łódzkie
małopolskie	podkarpackie	wielkopolskie
zachodniopomorskie	kujawsko-pomorskie	lubelskie
mazowieckie	opolskie	podlaskie
pomorskie	śląskie	świętokrzyskie
warmińskie		

Sprzedawca	
Międzynarodowy targ owocowy	
Pośrednictwo sprzedaży warzyw i owoców...	
Targ owocowy w Polsce, sieć hurtowni	
Targ rolny	

Rysunek 5.5. Zastosowanie fragmentatorów do filtrowania transakcji sprzedaży Firmy X

Źródło: Opracowanie własne.

Filtrowanie danych według formatu (autofiltr, opcja *Filtruj według koloru*) pozwala wybrać wartości o określonym kolorze czcionki, kolorze wypełnienia komórki lub też zawierające określoną ikonę komórki (wstawioną za pośrednictwem *Formatowania warunkowego*).

Opcje filtrowania według kryteriów zależą od formatu danych w danej kolumnie (tabela 5.1).

Tabela 5.1. Opcje filtrowania według kryteriów w zależności od typu danych

Filtr dat	Filtr liczb	Filtr tekstu
	Równa się, Filtr niestandardowy	
	Nie równa się	
	Między	
Większe niż, Większe niż lub równe, Mniejsze niż, Mniejsze niż lub równe, Pierwsze 10, Powyżej średniej, Poniżej średniej	Przed, Po, Jutro, Dzisiaj, Wczoraj, Następny tydzień, Ten tydzień, Ubiegły tydzień, Następny miesiąc, Ten miesiąc, Ubiegły miesiąc, Następny kwartał, Ten kwartał, Ubiegły kwartał, Następny rok, Ten rok, Ubiegły rok, Od początku roku, Wszystkie daty w okresie	Zaczyna się od, Kończy się na, Zawiera, Nie zawiera

Źródło: Opracowanie własne.

Za pośrednictwem opcji *Filtr niestandardowy* (dostępnej w przypadku każdego typu danych) można zdefiniować dla jednej kolumny dwa kryteria filtrowania

połączone operatorem LUB bądź ORAZ. W przypadku konieczności zdefiniowania większej liczby kryteriów należy zastosować filtr zaawansowany. Zaletą filtra zaawansowanego jest również możliwość wstawiania wyników filtrowania w inne miejsce aktywnego arkusza – autofiltr pozwala jedynie na ukrycie rekordów niespełniających kryteriów wyboru.

W Firmie X filtr zaawansowany wykorzystywany jest do weryfikowania poprawności i oczyszczania danych, a przede wszystkim do poprawiania błędów wynikających z braku słowników i jednolitego standardu wprowadzania danych, co znacznie obniża jakość danych i jest szczególnie częste w przypadku danych teleadresowych [Masewicz, 2009]. Jak wykazały analizy przeprowadzone z wykorzystaniem narzędzia *Suma częściowa*, w danych transakcji sprzedaży występują różne nazwy dla tych samych produktów (np. marchew i marchewka), co powoduje, że sumaryczne dane dotyczące sprzedaży są nieprawidłowe.

Aby w pełni wykorzystać możliwości filtra zaawansowanego, przy definiowaniu kryterium stosowane są odpowiednie znaki i symbole, których znaczenie i przykłady zastosowania opisano w tabeli 5.2.

Tabela 5.2. Przykłady wykorzystania filtra zaawansowanego do poprawy jakości danych

Znak/ wartość	Znaczenie	Przykład wykorzystania w Firmie X
= <i>"tekst"</i> lub <i>tekst</i>	Zwraca wartości zaczynające się od <i>tekst</i>	– ujednoczenie nazw produktów (np. marchew, marchewka, marchewki) – wyszukanie wszystkich rodzajów marchwi (np. młoda, świeża) celem ustandaryzowania nazewnictwa
= <i>"=tekst"</i>	Zwraca rekordy z wartością <i>tekst</i> w danej kolumnie (dokładne wyszukiwanie)	– po ujednoczeniu nazw produktów, sprawdzenie, czy liczba rekordów z dokładnym wyszukaniem prawidłowej nazwy zgadza się z liczbą rekordów wg wyszukiwania po początku nazwy
<> <i>tekst</i>	Zwraca wartości różne od <i>tekst</i>	– wyszukanie nieprawidłowych nazw
?	Zastępuje dowolny pojedynczy znak	– wyszukanie najczęściej występujących literówek w nazwach produktów
*	Zastępuje dowolną liczbę znaków	– wyszukanie grup produktów celem ujednoczenia ich nazw (np. młoda marchew, marchew młoda, młody ziemniak, ziemniak młody)
~ (tylda)	Wstawiony po niej znak (? , * , ~) traktowany jest jako zwykły znak	– wyszukanie nieprawidłowych znaków w nazwach produktów

Źródło: Opracowanie własne.

Zastosowanie filtra zaawansowanego wymaga zdefiniowania tzw. kryterium filtrowania. Jest to zakres komórek, który składa się z nazw kolumn (muszą być one umieszczone w pierwszym wierszu zakresu kryterium), według których dane będą filtrowane, oraz warunków, które w danych kolumnach powinny być

spełnione. Przykłady różnych kryteriów filtrowania, wyniki ich działania oraz ustawienie parametrów filtra zaawansowanego (na przykładzie ostatniego kryterium) prezentuje rysunek 5.6.

Produkt	Produkt	Produkt	Produkt	Produkt	Produkt	Produkt	Produkt	Produkt
marchew	=\"marchew\"	=\"marchew\"	=\"marchew\"	=\"mar?hew\"	mar?hew	=\"mar*?hew\"	=\"mar?hew\"	<?marchewka

Produkt	Produkt	Produkt	Produkt	Produkt	Produkt	Produkt	Produkt	Produkt
Marchew	Marchew	Marchew	Marchew	Marchew	Marchew	Marchew	Marchew	Marchew
Marchewka	Marchewka	Marchewka	Marchewka	Marchewka	Marchewka	Marchewka	Marchewka	Marchewka mloda
marchewka mloda	marchewka mloda	marchewka mloda	marchewka mloda	marchewka mloda	marchewka mloda	marchewka mloda	marchewka mloda	Mloda marchewka
marchew	marchew	marchew	Mloda marchewka	marchew	Marchewka	Marchewka	Marchewka	Marchewka
			marchew	Marchewka	Marchewka	Marchewka	Marchewka	Marchewka
				Mar?hewka	Mar?hewka	Mar?hewka	Mar?hewka	Mar?hewka

Rysunek 5.6. Przykłady wykorzystania filtra zaawansowanego do oczyszczania danych

Źródło: Opracowanie własne.

W filtrze zaawansowanym można stosować dowolną liczbę warunków w ramach jednego kryterium filtrowania, dzięki czemu oczyszczanie danych nie musi być żmudnym procesem. Można przykładowo w ramach jednej iteracji wyszukać różne błędne nazwy, wprowadzając poszczególne warianty w kolejnych wierszach kryterium filtrowania.

5.3 Przypadek 3 – filtrowanie danych z zastosowaniem formuł w analizie importu i eksportu towarów

Firma X analizuje wartość importu i eksportu owoców z lat 2013 i 2014 celem określenia kierunków rozszerzenia listy potencjalnych krajów zbytu dla swoich produktów. Dane pobrane zostały z bazy Hinex udostępnionej na stronach GUS (rozdział 1, przypadek 1), a następnie wstępnie oczyszczone i przygotowane do analizy (rozdział 2, przypadek 1). Oczyszczona lista (plik *r5_import_i_eksport_filtr.xlsx*) zawiera jednak dublujące się dane – raz wartości importu/eksportu pojedynczych owoców, raz zbiorczo zsumowane wszystkie owoce. Żeby przy dalszych analizach uniknąć błędów wynikających z podwójnego liczenia kwot, z listy należy usunąć wartości zbiorcze. Można to wykonać ręcznie (porównaj rozdział 2, przypadek 1), bądź zautomatyzować proces, stosując filtr zaawansowany.

Jak łatwo zauważyć (arkusz *EKSPORT_2014*), wartości zbiorcze nie zawierają danych w kolumnie ilość, co oznacza, że może być ona wykorzystana jako kolumna

filtrowania. Wystarczy zatem wybrać rekordy, w których wartość w kolumnie ilość spełnia warunek: >0 i zbudować z nich osobną listę, będącą podstawą dalszych analiz. Wynik filtrowania dla danych dotyczących eksportu w 2014 roku prezentuje arkusz *EKSPORT_2014_o*.

Zdarza się, że przy czyszczeniu danych pewne istotne informacje zostają usunięte. W omawianym przypadku podczas usuwania wierszy podsumowujących „zginęła” informacja o kategorii poszczególnych krajów (zawarta tylko w tych wierszach). Żeby nie utracić tej informacji, należało wcześniej przygotować arkusz *Kategorie krajów*. W tym celu skopiowano z arkusza wejściowego (*EKSPORT_2014*) do nowego potrzebne kolumny (*kraj*, *kod kraju*) oraz do nowej kolumny (*kategoria*) wklejono odpowiednie nazwy kategorii, a następnie usunięto zbędne wiersze (puste oraz podsumowujące). Do tego celu wykorzystano filtr zaawansowany (rysunek 5.7).

Lista wyjściowa:			Kryterium filtrowania:		
kraj	kategoria	kod kraju	kraj	kategoria	kod kraju
Białoruś	Kraje Eur. Śr.-Wsch.	BY			
Moldawia	Kraje Eur. Śr.-Wsch.	MD			
Rosja	Kraje Eur. Śr.-Wsch.	RU			
Ukraina	Kraje Eur. Śr.-Wsch.	UA			
RAZEM Kraje rozwijające się :			Lista wynikowa		
			kraj	kategoria	kod kraju
			Białoruś	Kraje Eur. Śr.-W:	BY
			Moldawia	Kraje Eur. Śr.-W:	MD
			Rosja	Kraje Eur. Śr.-W:	RU
			Ukraina	Kraje Eur. Śr.-W:	UA
			Algeria	Kraje rozwijając	DZ
			Angola	Kraje rozwijając	AO
			Antarktyda	Kraje rozwijając	AQ
			Arabia Saudyjska	Kraje rozwijając	SA
			Armenia	Kraje rozwijając	AM
			Azerbejdżan	Kraje rozwijając	AZ

Rysunek 5.7. Przykłady wykorzystania filtra zaawansowanego do oczyszczania danych

Źródło: Opracowanie własne.

Przy tworzeniu kryterium filtrowania użyto formuły logicznej (dającej wynik PRAWDA lub FAŁSZ) sprawdzającej, czy komórka zawierająca kategorię kraju nie jest pusta ($<>\"$). Dodatkowo zaznaczono opcję *Tylko unikatowe rekordy*, ponieważ dane na nieoczyszczonej liście powtarzają się.

Aby móc ocenić wielkość eksportu w obrębie poszczególnych kategorii krajów oraz kontynentów, w arkuszu z wartościami eksportu dodano odpowiednie kolumny i za pośrednictwem funkcji WYSZUKAJ.PIONOWO (opisanej w rozdziale 3) dla każdego rekordu pobrano stosowne informacje z arkuszy pomocniczych (*Kraje i kontynenty* oraz *Kategorie krajów*).

Tak przygotowany arkusz z danymi dotyczącymi eksportu w 2014 roku posłużył do wstępnej selekcji krajów jako potencjalnych kierunków eksportu

poszczególnych owoców. Postanowiono wybrać te kraje, do których eksport był ilościowo lub wartościowo większy od średniej w ramach danego produktu. Osobno rozpatrzono kraje Unii Europejskiej i spoza Unii. W pierwszej kolejności za pośrednictwem filtra zaawansowanego wybrano kraje, które są największymi odbiorcami gruszek świeżych. W tym celu zdefiniowano kryterium filtrowania z czterema opcjonalnymi warunkami (połączonych operatorem LUB), z których każdy składał się z trzech warunków obligatoryjnych (połączonych operatorem ORAZ). W pierwszym przypadku poszczególne kryteria muszą znajdować się w różnych wierszach, w drugim przypadku – w jednym wierszu (rysunek 5.8).

A	B	C	D	E	H	I	L	M	N	O	P
EKSPORT - rok 2014, 4 kwartały											
Kod	Nazwa towaru	Kraj	Grupa krajów	Większe od średniej	Podział formuły						
	Gruszek świeże		<Unia Europejska	PRAWDA	=I9>ŚREDNIA.WARUNKÓW(\$H\$9:\$H\$119;\$D\$9:\$D\$5119;\$D\$3:\$D\$5119;\$B\$3)						
	Gruszek świeże		<Unia Europejska	PRAWDA	=I9>ŚREDNIA.WARUNKÓW(\$I\$9:\$I\$119;\$D\$9:\$D\$5119;\$D\$3:\$D\$5119;\$B\$3)						
	Gruszek świeże		Unia Europejska	PRAWDA	=I9>ŚREDNIA.WARUNKÓW(\$H\$9:\$H\$119;\$D\$9:\$D\$5119;\$D\$5:\$D\$5119;\$B\$3)						
	Gruszek świeże		Unia Europejska	PRAWDA	=I9>ŚREDNIA.WARUNKÓW(\$I\$9:\$I\$119;\$D\$9:\$D\$5119;\$D\$5:\$D\$5119;\$B\$3)						
Lista danych:						Wynik filtrowania:					
Kod	Nazwa towaru	Kraj	Grupa krajów	Kontynent	Ilość	Wartość w zł	Kod	Nazwa towaru	Kraj	Grupa krajów	
080830	Gruszek świeże	Białoruś	Kraje Eur. Śr.-Wsch. Europa		20798518	36 420 997 zł	080830	Gruszek świeże	Białoruś	Kraje Eur. Śr.-Wsch.	
080830	Gruszek świeże	Rosja	Kraje Eur. Śr.-Wsch. Azja		7278082	18 797 720 zł	080830	Gruszek świeże	Rosja	Kraje Eur. Śr.-Wsch.	
080830	Gruszek świeże	Ukraina	Kraje Eur. Śr.-Wsch. Europa		3349327	9 396 981 zł	080830	Gruszek świeże	Ukraina	Kraje Eur. Śr.-Wsch.	
080830	Gruszek świeże	Armenia	Kraje rozwijające si Azja		100512	215 653 zł	080830	Gruszek świeże	Belgia	Unia Europejska	
080830	Gruszek świeże	Bośnia i Hercegowina	Kraje rozwijające si Europa		81264	85 351 zł	080830	Gruszek świeże	Cypr	Unia Europejska	
080830	Gruszek świeże	Czarnogóra	Kraje rozwijające si Europa		22560	35 351 zł	080830	Gruszek świeże	Litwa	Unia Europejska	
080830	Gruszek świeże	Kajmany	Kraje rozwijające si Ameryka Środkowa		3072	12 876 zł	080830	Gruszek świeże	Łotwa	Unia Europejska	
080830	Gruszek świeże	Kazachstan	Kraje rozwijające si Azja		1344919	2 643 760 zł	080830	Gruszek świeże	Wielka Brytania	Unia Europejska	
080830	Gruszek świeże	Kosowo	Kraje rozwijające si Europa		14820	42 497 zł					

Rysunek 5.8. Wykorzystanie filtra zaawansowanego do filtrowania według wielu kryteriów

Źródło: Opracowanie własne.

W ramach każdej z czterech grup warunków opcjonalnych określono trzy warunki łączne w postaci: rodzaj produktu, grupa krajów (kraj z UE i spoza) oraz ilość/wartość eksportu przekraczająca średnią dla danego produktu i danej grupy krajów (co zdefiniowano, wykorzystując funkcję ŚREDNIA.JEŻELI). Wyniki filtrowania zaprezentowano na rysunku 5.8. Podobne czynności powtórzono dla pozostałych produktów. Tym samym uzyskano listę krajów, do których eksport poszczególnych owoców będzie potencjalnie najkorzystniejszy. Wyniki zaprezentowano w arkuszu *eksport_lista krajów*.

5.4. Przypadek 4 – tabele przestawne w analizie sprzedaży produktów

Firma X na bieżąco musi monitorować i analizować sprzedaż swoich produktów, żeby prawidłowo i szybko reagować na zmiany w popycie na rynku. Do tego celu firma wykorzystuje tabele przestawne, które są użytecznym narzędziem analizy dużych zbiorów danych, jakim są dane sprzedaży [Próchnicki, 2014]. Tabela przestawna (ang. *pivot table*) jest obok wykresów jednym z podstawowych elementów

kokpitu menedżerskiego [Walkenbach, Aleksander, 2014], którego podstawową rolą jest dostarczanie właściwym użytkownikom, we właściwym czasie, właściwych informacji w celu podejmowania lepszych decyzji ([Dudycz, 2010 za: Ziora, 2012]) oraz zwiększenia wydajności działania organizacji [Ziuziański, Furmaniewicz, 2015]. Tworzenie kokpitów (pulpitów) menedżerskich w MS Excel zostanie omówione w rozdziale 7.

Tabela przestawna jest narzędziem pozwalającym na utworzenie interaktywnego widoku (określanego jako raport tabeli przestawnej) na podstawie danych źródłowych – bazy danych umieszczonej w arkuszu lub w pliku zewnętrznym [Aleksander i in., 2015]. Jest to dynamiczne zestawienie danych, które umożliwia przekształcanie niekończących się ciągów rzędów i kolumn w prezentacje danych o interesujących użytkownika układzie, stopniu szczegółowości i zakresie danych [Walkenbach, 2013].

Tabela przestawna składa się z czterech obszarów: Wartości, Wiersze, Kolumny i Filtry, w których można dowolnie umieszczać pola z listy źródłowej, definiując w ten sposób pożądany układ tabeli. Wartości w polach tabeli można dowolnie sortować, filtrować czy też grupować. Tabele przestawne wykorzystywane są do analizy danych w różnych obszarach działania przedsiębiorstwa, a przede wszystkim w sprzedaży, gdzie liczba transakcji i ich szczegółów są duże, a ich eksploracja pozwala pozyskać wiedzę umożliwiającą między innymi skutecznie zarządzać relacjami z klientami, budować oczekiwaną strukturę asortymentową czy też wyznaczać kierunki rozwoju przedsiębiorstwa.

Firma X wykorzystuje tabele przestawne do analizy i monitorowania danych sprzedaży (plik *r5_faktury_sprzedazy_tabele.xlsx*, arkusz *faktury sprzedaży*) w następujących przekrojach:

- według grup produktów i produktów,
- według klientów,
- według sprzedawców,
- według jednostek terytorialnych,
- analiza w czasie (zaprezentowana w ramach przypadku 5).

W tabeli wspierającej analizę według produktów (rysunek 5.9) zestawiono najważniejsze wskaźniki dotyczące sprzedaży: ilość, wartość przychodu ze sprzedaży oraz zysk. W obszarze Filtrów umieszczono pole *Kategoria produktu*. W wierszach umieszczono listę produktów, stosując przy tym następującą hierarchię: *Rodzaj produktu*, *Jednostka miary* (ponieważ ilość w przypadku owoców prezentowana jest w kilogramach lub w sztukach, więc nie można jej sumować) oraz konkretny *Produkt*. Aby utworzyć pole *Rodzaj produktu* (podział na Owoce krajowe, Owoce importowane oraz Warzywa), należało, przed dodaniem pola *Jednostka miary*, posortować produkty według listy niestandardowej, którą utworzono w ramach przypadku 1 (Opcja *Sortuj*, Więcej opcji sortowania, Rosnąco według Produktu, Więcej

opcji, Kolejność sortowania pierwszego klucza – por. rysunek 5.9), a następnie pogrupować odpowiednie produkty (opcja *Grupuj*).

The screenshot displays a pivot table with the following data:

Kategoria produktu	(Wszystko)			
Grupy produktów/ produkty	Suma z Ilość	Przychód	Zysk	
Owoce krajowe				
kg	57 795	281 973 zł	105 827 zł	
Jabłka świeże	17 589	33 090 zł	26 055 zł	
Gruszkki świeże	27 072	102 363 zł	68 094 zł	
Truskawki	13 134	146 520 zł	11 679 zł	
Owoce importowane				
kg	104 284	545 375 zł	133 170 zł	
Winogrona	38 343	293 864 zł	17 139 zł	
Pomarańcza	38 232	162 306 zł	82 365 zł	
Banany	27 719	89 104 zł	33 866 zł	
szt	77 376	200 804 zł	100 541 zł	
Brzoskwinie	38 907	91 592 zł	45 185 zł	
Nektaryniki	38 469	109 212 zł	55 356 zł	
Warzywa				
kg	103 248	310 032 zł	65 239 zł	
Marchew	38 680	54 292 zł	14 932 zł	
Pomidor	38 149	180 548 zł	27 952 zł	
Ziemniaki	26 419	75 193 zł	22 355 zł	
Suma końcowa	342 713	1 338 184 zł	404 777 zł	

Overlaid on the table are two dialog boxes:

- Pola tabeli przestawnej**: A dialog for selecting fields to be placed in the pivot table. It includes a list of fields (Data, Produkt, etc.) and options for where to place them (Rows, Columns, Values).
- Sortowanie (Produkt)**: A dialog for setting sorting options. It shows 'Sortuj pole Produkt w kolejności' and 'Ważność sortowania pierwszego klucza' set to 'Jabłka świeże, Gruszkki świeże, Truskawki, Winogrona'.

Rysunek 5.9. Tabela przestawna wspierająca analizę sprzedaży według produktów (arkusz *analiza wg produktów*)

Źródło: Opracowanie własne.

Po określeniu właściwej struktury tabeli zdefiniowano odpowiednie formatowanie wartości liczbowych, co poprawiło czytelność danych. Wybór odpowiedniej formy prezentacji tabeli (*forma kompaktowa* – zaprezentowana na rysunku 5.9, *konspektu, tabelaryczna*) wpłynął na szybkość odczytu zawartych w danych informacji.

W ramach analizy danych według klientów Firma X wykorzystuje tabelę przestawną prezentującą ranking odbiorców przynoszących najwyższe zyski (arkusz *analiza wg klientów*). Układ tabeli widoczny jest na rysunku 5.10. Żeby lepiej ocenić udział poszczególnych klientów w generowaniu zysków, stosowne wartości bezwzględne zmieniono na wartości procentowe (opcja *Ustawienia pola wartości – % wartości kolumny*). Następnie dla pola *Klient* zdefiniowano odpowiedni filtr pozwalający wybrać siedmiu najlepszych klientów (*Filtry wartości, Pierwsze 10*). Dla zwiększenia czytelności danych wartości zysku zostały posortowane według sumy końcowej. Zastosowano również formatowanie warunkowe dla wyróżnienia wartości powyżej średniej (dla każdej z kolumn zdefiniowana została osobna reguła z wykorzystaniem *Malarza formatów*).

Kolejny rodzaj analizy realizowany z wykorzystaniem tabel przestawnych to monitoring i ocena pracy sprzedawców (arkusz *analiza wg sprzedawców*). Firma X prowadzi sprzedaż hurtową owoców i warzyw za pośrednictwem czterech dystrybutorów/pośredników, którzy posiadają punkty dystrybucji na terenie całej Polski.

Województwo		(Wszystko)		
Miejscowość		(Wszystko)		
Suma z Zysk		kategoria produktu		Suma
Klient		Owoce	Warzywa	końcowa
Zywność Gdańska Sp. z o.o.		21,67%	5,39%	19,81%
Owoce i warzywa Sp. z o.o.		16,65%	27,98%	17,95%
ABC Sp. z o.o.		15,66%	10,79%	15,10%
Owoce z Zachodniego Pomorza, Sp.		10,13%	31,46%	12,56%
Hurtownia spożywcza w Białymstoku		12,93%	8,76%	12,45%
Zdrowa żywność SA		11,97%	5,58%	11,24%
Zakłady spożywcze SA		10,99%	10,03%	10,88%
Suma końcowa		100,00%	100,00%	100,00%

FILTRY: Województwo, Miejscowość

KOLUMNY: Kategorie produktu

WIERSZE: Klient

WARTOŚCI: Suma z Zysk

Rysunek 5.10. Tabela przestawna prezentująca ranking klientów przynoszących największe zyski (arkusz *analiza wg klientów*)

Źródło: Opracowanie własne.

Niewielka liczba sprzedawców sprawia, że bardzo ważne są bieżące monitorowanie i ocena efektów ich pracy oraz kształtowanie na tej podstawie odpowiedniej polityki współpracy ze sprzedawcami. Jednym ze sposobów realizacji polityki sprzedaży/dystrybucji jest sposób motywowania sprzedawców poprzez wprowadzenie prowizji dla nich w zależności od wygenerowanego przez nich zysku.

Do analizy i oceny poszczególnych sprzedawców wykorzystano tabelę przestawną, prezentującą generowane przez nich zyski. Dodatkowo zdefiniowano dwa pola obliczeniowe (wstążka *Analiza*, opcja *Pola, elementy i zestawy*): Prowizja (5% z Zysku) oraz Zysk po wypłaceniu prowizji (rysunek 5.11).

Lata	2015		
Kwartaly	Kwartał2		
	Suma z Zysk	Suma z Prowizja dla sprzedawcy	Suma z Zysk po prowizji
Etykiety wierszy	41 378 zł	2 069 zł	39 309 zł
☐ Pośrednictwo sprzedaży warzyw i ow			
Owoce	33 739 zł	1 687 zł	32 052 zł
Warzywa	7 639 zł	382 zł	7 257 zł
☐ Targ owocowy w Polsce, sieć hurtow	14 558 zł	728 zł	13 830 zł
Owoce	14 558 zł	728 zł	13 830 zł
☐ Międzynarodowy targ owocowy	5 742 zł	287 zł	5 455 zł
Owoce	5 742 zł	287 zł	5 455 zł
☐ Targ rolny	3 485 zł	174 zł	3 310 zł
Warzywa	3 485 zł	174 zł	3 310 zł
Suma końcowa	65 162 zł	3 258 zł	61 904 zł

Wstawianie pola

Nazwa: Prowizja dla sprzedawcy
Formuła: = Zysk*5%

Wstawianie pola obliczeniowego

Nazwa: Zysk po prowizji
Formuła: = Zysk - Prowizja dla sprzedawcy

Pola: Termin zapłaty, Kwota zapłaty, Sposób zapłaty, Przychód, Zysk, Produkt2, Kwartały, Lata

Rysunek 5.11. Tabela przestawna obliczająca prowizję kwartalną dla sprzedawców (arkusz *analiza wg sprzedawców*)

Źródło: Opracowanie własne.

Ponieważ prowizja wypłacana jest w cyklu kwartalnym, w tabeli przestawnej w obszarze filtra umieszczono dodatkowe pola: *Lata* i *Kwartaly*, dzięki którym można nie tylko obliczać prowizję kwartalną, ale również porównywać dane w układzie kwartalnym z wartościami poprzednich okresów czy też adekwatnych okresów z poprzednich lat. Ponieważ w danych źródłowych jest tylko pole *Data*

(ważne jest, żeby format daty był prawidłowo zdefiniowany), w pierwszej kolejności umieszczono je w obszarze wierszy tabeli. To skutkowało automatycznym pogrupowaniem danych według różnych jednostek czasu (m.in. rok, kwartał, miesiąc) i dodaniem odpowiednich pól do listy pól tabeli przestawnej, które następnie można przenieść do dowolnego obszaru, a w tym przypadku do obszaru filtra.

Ocena sprzedaży według jednostek terytorialnych (arkusz *analiza wg obszarów*) to kolejny ważny obszar analiz prowadzonych przez Firmę X, których celem jest określenie kierunków inwestycji i rozwoju poszczególnych regionów Polski. Przy tworzeniu hierarchii jednostek terytorialnych wzorowano się na nomenklaturze NTS (Nomenklatura Jednostek Terytorialnych do Celów Statystycznych) opracowanej przez GUS [GUS, 2015] i wyróżniono dwa pierwsze (z pięciu) poziomy: nieadministracyjny Regiony (NTS1) i administracyjny Województwa (NTS2). Do wyodrębnienia regionów w tabeli przestawnej wykorzystano polecenie *Element obliczeniowy...* (wstążka *Analiza*, opcja *Pola, elementy i zestawy*) dla pola Województwo (rysunek 5.12).

The screenshot shows an Excel pivot table with the following data:

Kategoria produktu	(Wszystko)		
Produkt	(Wszystko)		
	Suma z	Suma z	
	Przychód	Zysk	
Regiony			
pomorskie	145 428 zł	45 126 zł	
warmińsko-mazurskie	76 910 zł	15 454 zł	
kujawsko-pomorskie	129 798 zł	40 887 zł	
Region północny	352 136 zł	101 467 zł	
podlaskie	166 310 zł	28 367 zł	
lubelskie	75 206 zł	23 017 zł	
podkarpackie	58 294 zł	13 769 zł	
świętokrzyskie	73 580 zł	22 630 zł	
Region wschodni	373 390 zł	87 782 zł	
mazowieckie	51 547 zł	25 614 zł	
łódzkie	77 732 zł	20 900 zł	
Region centralny	129 280 zł	46 513 zł	
zachodniopomorskie	71 261 zł	28 621 zł	
lubuskie	56 016 zł	19 548 zł	
wielkopolskie	66 250 zł	19 022 zł	
Region północno-zachodni	193 526 zł	67 191 zł	
dolnośląskie	71 959 zł	34 410 zł	
opolskie	73 936 zł	19 664 zł	
Region południowo-zachodni	145 894 zł	54 074 zł	
małopolskie	81 144 zł	24 789 zł	
śląskie	62 813 zł	22 960 zł	
Region południowy	143 958 zł	47 750 zł	
Razem	1 338 184 zł	404 777 zł	

The dialog box "Wstaw element obliczeniowy w 'Województwo'" is open, showing the following details:

- Nazwa: Region północny
- Formuła: =pomorskie + 'warmińsko-mazurskie' + 'kujawsko-p
- Pola: Rabat, Ilość, Jednostka miary, Wartość zakupu, Sprzedawca, Klient, Miejscowość, Województwo
- Elementy: pomorskie, warmińsko-mazurskie, kujawsko-pomorskie, Region północny, podlaskie, lubelskie, podkarpackie, świętokrzyskie

Rysunek 5.12. Ocena przychodów i zysków według regionów z wykorzystaniem tabeli przestawnej (arkusz *analiza wg obszarów*)

Źródło: Opracowanie własne.

W pierwszej kolejności odpowiednio zdefiniowano wszystkie sześć regionów, a następnie uporządkowano wiersze w tabeli przestawnej, tak żeby nowe elementy – *Regiony* – widoczne były jako podsumowanie właściwych województw. Ponieważ dodanie regionów spowodowało podwojenie wartości sumy końcowej, utworzono dodatkowy element obliczeniowy *Razem* (jako suma poszczególnych regionów) oraz wyłączono sumy końcowe dla wierszy (wstążka *Projektowanie*,

opcja *Sumy końcowe*). Utworzone regiony widoczne będą we wszystkich tabelach przestawnych opartych na tej samej liście źródłowej, jednak nie zostanie zachowane ich sortowanie, warto zatem zachować kolejność sortowania w postaci *Listy niestandardowej*.

W efekcie wykonanych działań można stwierdzić, że najwyższe przychody ze sprzedaży owoców i warzyw w perspektywie ostatnich 18 miesięcy Firma X osiąga z regionu wschodniego, natomiast najwyższe zyski z regionu północnego. Najmniej przychodów i zysków generuje region centralny. Dalsza analiza powinna pomóc określić kryteria wpływające na najlepsze i najgorsze wyniki oraz wspomóc proces formułowania strategii rozwoju przedsiębiorstwa w zakresie produktów, cen i promocji.

Elementy obliczeniowe można wykorzystać nie tylko do sumowania wartości, ale również do wyliczania różnicy pomiędzy wartościami czy też innych obliczeń. Przykładowo, jeżeli na liście źródłowej istnieje kolumna o nazwie Miesiąc, można obliczyć zmianę wartości (np. ilości, zysku czy przychodu) w stosunku do poprzedniego miesiąca.

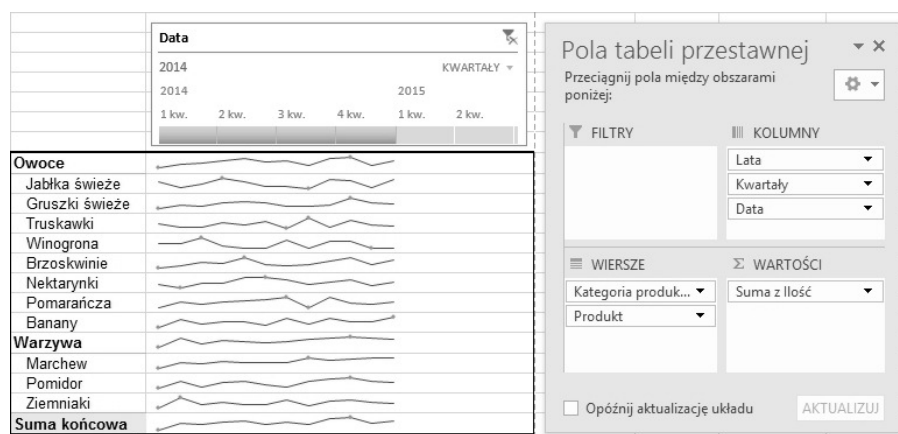
Utworzenie *Elementu obliczeniowego* niesie jednak za sobą ograniczenia, a mianowicie skutkuje brakiem możliwości grupowania danych we wszystkich tabelach przestawnych bazujących na tej samej liście danych (identycznym zakresie źródłowym). Nie można też tworzyć *Elementów obliczeniowych* w tabelach, w których dokonano wcześniej ręcznego grupowania danych. Dotyczy to również automatycznego grupowania według dat, co znacznie utrudnia analizę danych w czasie. W przypadku gdy ma być ona przeprowadzona, należy zrezygnować z *Elementów obliczeniowych* na rzecz ręcznego grupowania wartości.

5.5. Przypadek 5 – wykorzystanie fragmentatorów i osi czasu do filtrowania oraz integracji tabel i wykresów przestawnych

Sprzedaż owoców i warzyw charakteryzuje duża sezonowość. Z perspektywy zarządzającego nie jest to pozytywne zjawisko i wymaga podjęcia działań, które będą niwelować skutki wahań sezonowych. Firma X na bieżąco prowadzi analizę trendu sprzedaży (dane źródłowe zawiera plik *r5_faktury_sprzedazy_tabele.xlsx*, arkusz *faktury sprzedazy*), wykorzystując w tym celu, bazujące na tabelach przestawnych, wykresy przebiegu w czasie (ang. *sparklines*) oraz tworzone za pośrednictwem dodatku Power View (por. rozdział 4) wykresy panelowe. Wykres panelowy (ang. *panel chart*) jest zestawem kilku małych wykresów tego samego typu, przedstawiających tą samą zmienną (np. przychód ze sprzedaży produktów),

zróznicowaną pod względem wybranej cechy jakościowej (np. sprzedawcy, klienta) [http://www.contextures.com/excelpanelchart.html, 2015]. Wykres panelowy ułatwia zrozumienie wielowymiarowych danych [http://skuteczneraporty.pl/blog/wielokrotnosc-czyli-wykres-panelowy/, 2015].

Utworzone na bazie tabeli przestawnej wykresy przebiegu w czasie (wstążka *Wstawianie*) pokazują wahania w ilości sprzedaży poszczególnych owoców i warzyw w 2014 roku (rysunek 5.13).



Rysunek 5.13. Wykorzystanie tabel przestawnych oraz wykresów przebiegu w czasie do analizy wahań sezonowych w ilości sprzedaży produktów (arkusz *analiza w czasie*)

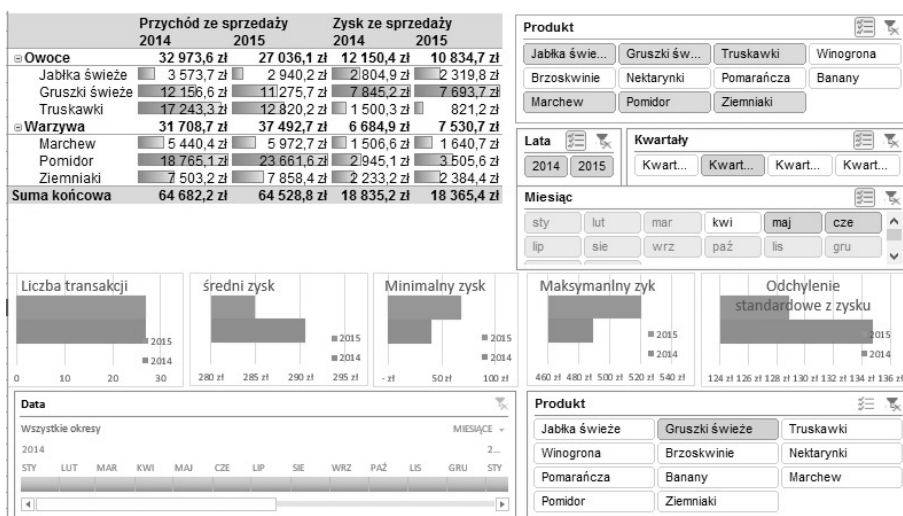
Źródło: Opracowanie własne.

Okres czasu został zawężony za pośrednictwem osi czasu (wstążka *Wstawianie*), dzięki której można wybierać dowolne jednostki (lata, kwartały, miesiące czy dni) i przedziały czasowe do analizy. Jak widać, Firma X dobrze sobie radzi z sezonowością. Wprawdzie wahania sprzedaży poszczególnych produktów występują, jednak są one wzajemnie kompensowane.

Ponieważ w przypadku większości produktów oferowanych przez Firmę X nie można całkowicie wyeliminować sezonowości, dlatego też monitoruje ona sprzedaż, porównując aktualne wartości przychodów i zysków z adekwatnymi okresami w poprzednich latach. Wykorzystanie tabel i wykresów przestawnych pozwala na dokonywanie analizy bez konieczności wykonywania dodatkowych obliczeń i jednocześnie umożliwia wizualizację efektów porównania, co znacznie automatyzuje i przyspiesza cały proces.

Tabela przestawna na rysunku 5.14 prezentuje porównanie przychodów i zysków ze sprzedaży poszczególnych produktów w latach 2014–15. Aby zwiększyć elastyczność analizy, dodano fragmentatory (wstążka *Wstawianie*, opcja *Fragmentator* lub polecenie *Dodaj jako fragmentator* dostępne na liście pól tabeli przestawnej

w ramach menu podręcznego każdego z pól) umożliwiające filtrowanie danych według produktów oraz okresów (lat, kwartałów i miesięcy). Ponieważ na liście pól nie były dostępne poszczególne okresy (nie ma stosownych kolumn w źródłowej liście danych), do tabeli wstawiono pole *Data* i po rozpoznaniu formatu daty, MS Excel automatycznie utworzył odpowiednią hierarchię okresów. Wówczas kwartały i miesiące usunięto z tabeli, pozostały one jednak na liście pól i można je było wykorzystać do utworzenia fragmentatora (dla miesiąca można go utworzyć na podstawie pola *Data*). Ustawienie we fragmentatorach opcji *Wybór wielokrotny* zwiększyło elastyczność w zakresie wyboru analizowanych okresów czy też produktów.



Rysunek 5.14. Wykorzystanie fragmentatorów i osi czasu do filtrowania danych i łączenia tabel/wykresów przestawnych (arkusz *analiza w czasie*)

Źródło: Opracowanie własne.

Z danych zaprezentowanych w tabeli przestawnej na rysunku 5.14 wynika, że w maju i czerwcu 2015 roku zarówno przychody, jak i zyski ze sprzedaży owoców krajowych w stosunku do roku poprzedniego spadły, natomiast wzrosły przychody i zyski ze sprzedaży warzyw. W rezultacie przychody i zyski ze sprzedaży produktów krajowych w nieznacznym stopniu spadły. Dzięki zastosowaniu formatowania warunkowego (*Pasków danych*), łatwo zauważyć, które produkty mają duży udział w generowaniu przychodów (pomidory) i zysków (gruszki), a które mały (odpowiednio jabłka oraz truskawki i marchew).

Aby zrozumieć przyczyny zmian, konieczna była dodatkowa analiza na podstawie takich wskaźników, jak liczba transakcji, średnia, minimalna i maksymalna wartość czy też odchylenie standardowe, obejmująca poszczególne produkty.

Umieszczenie wielu elementów w jednej tabeli znacznie zmniejsza jej czytelność, z drugiej strony konieczna jest analiza wielu parametrów jednocześnie. Rozwiązaniem było przygotowanie kilku wykresów przestawnych połączonych fragmentatorami (rysunek 5.14). Opcja *Połączenia raportu* pozwoliła zdefiniować, które tabele/wykresy, bazujące na tym samym źródle danych, będą filtrowane za pośrednictwem danego fragmentatora. Ponieważ wykres przestawny powiązany jest z tabelą, na podstawie której jest tworzony (zmiana w tabeli powoduje zmianę na wykresie i odwrotnie), każdy z wykresów przygotowany został wraz z oddzielną kopią tabeli przestawnej. Dzięki kopiowaniu tabeli przestawnej wraz z wykresem (kopiowane są również połączenia z fragmentatorem), a następnie zmianie *Ustawień pola wartości* (na *Średnia*, *Minimum* itd.) oraz zastosowaniu szablonów wykresów, proces tworzenia poszczególnych wykresów został znacznie zautomatyzowany.

Utworzone wykresy podłączono do fragmentatorów tabeli przestawnej (*Lata*, *Kwartaly* i *Miesiąc*) oraz utworzono nowy (niezwiązany z tabelą) fragmentator – *Produkt* – pozwalający na wybór określonego towaru do analizy. Dodatkowo dodano oś czasu (połączoną ze wszystkimi wykresami), która umożliwi dokonywanie oceny poszczególnych wskaźników w konkretnym okresie czasu. Aby zwiększyć czytelność stworzonego zestawu wskaźników, na wykresach nie są wyświetlane przyciski pól (opcja *Ukryj wszystkie przyciski pól na wykresie*) oraz zmieniono ustawienia fragmentatorów (opcja *Ukryj elementy bez danych*) oraz liczbę wyświetlanych kolumn (*Rozmiar i właściwości*, grupa opcji *Położenie i układ*, *Liczba kolumn*).

Tabele przestawne można tworzyć na podstawie wielu lokalizacji danych, ale wymagane jest, żeby dane w każdej z nich miały tę samą strukturę i nagłówki kolumn. W przypadku gdy te warunki nie są spełnione, do utworzenia tabeli przestawnej na podstawie kilku lokalizacji konieczne jest wykorzystanie *Modelu danych*. Umożliwia on łączenie danych z wielu arkuszy oraz danych pochodzących z innych źródeł niż arkusz kalkulacyjny MS Excel [Winston, 2014]. Tworzenie modelu danych jest możliwe za pośrednictwem dodatku Power Pivot, co opisano w rozdziale 6.

Efekty zrealizowanych w ramach niniejszego studium przypadku analiz zaprezentowane zostały w arkuszu *analiza w czasie* (plik *r5_faktury_sprzedazy_tabele.xlsx*).

5.6. Możliwe zastosowania arkusza MS Excel w formie zadań

W celu zobrazowania możliwości zastosowania sum częściowych oraz tabel i wykresów przestawnych dostępnych w MS Excel można posłużyć się poniższymi zadaniami:

1. Do wstępnego porównania przychodów i zysków ze sprzedaży osiągniętych przez Firmę X w poszczególnych latach i miesiącach, w rozbiciu na kategorie produktów, należy wykorzystać sumy częściowe.
2. Firma X chce dokonać prognozy sprzedaży (ilościowo i wartościowo) na rok 2016. Zakłada, że sprzedaż owoców i warzyw wzrośnie o 5%, a ceny o 3% w stosunku do roku 2014. Dane należy zaprezentować w postaci wykresu przestawnego.
3. Firma X chce przeprowadzić analizę ilościową i wartościową sprzedaży poszczególnych produktów według klientów i okresów czasu (kwartałów). Należy stworzyć odpowiednią tabelę przestawną, której dane będą filtrowane za pośrednictwem fragmentatora i osi czasu. Dla zwiększenia czytelności wyników zastosować formatowanie warunkowe, zaznaczając i wyróżniając wartości powyżej średniej.
4. Firma X chce ocenić tendencje w zakresie eksportu i importu poszczególnych owoców i warzyw, żeby dokonać planów zakupów i sprzedaży na kolejny okres. Wykorzystując wykresy przestawne, należy porównać (obliczyć zmianę) wartości eksportu i importu owoców i warzyw do poszczególnych krajów w latach 2013–2014.
5. Firma X chce rozszerzyć swoją działalność eksportową, oferując produkty do przechowywania warzyw i owoców do pięciu krajów, do których eksport tego rodzaju towarów jest najwyższy. Na podstawie zbioru danych przygotowanego w ramach rozdziału 1 (zad. 3) należy dokonać, z wykorzystaniem tabel przestawnych, analizy porównawczej wielkości eksportu dla 5 ostatnich lat referencyjnych.

Zakończenie

Tabele i wykresy przestawne oraz sortowanie i filtrowanie danych to podstawowe narzędzia analityczne wykorzystywane przez firmy zarówno w systemach ewidencyjnych, jak i typowych rozwiązaniach analitycznych, w tym arkuszach kalkulacyjnych. Są one podstawowymi elementami kokpitów menedżerskich, których możliwości jako podstawowego narzędzia dostarczającego informacje niezbędne do sprawnego podejmowania decyzji zaprezentowane zostały w rozdziale 7.

Jak pokazano w niniejszym rozdziale, tabele i wykresy przestawne umożliwiają szybką analizę dużych zbiorów danych w różnych przekrojach, stosownie do potrzeb użytkownika. Dobór i zmiana danych do analizy możliwe są na każdym etapie, dzięki wbudowanym mechanizmom zaawansowanego filtrowania oraz fragmentatorom i osiom czasu. Możliwość zastosowania formatowania warunkowego oraz mnogość rodzajów wykresów pozwalają z kolei na dobór takich form prezentacji danych, które ułatwiają ich szybką i prawidłową interpretację. Interaktywność narzędzi, duże możliwości w zakresie personalizowania analiz połączone z intuicyjnością omawianych narzędzi sprawiają, że znajdują one coraz szersze i powszechniejsze zastosowanie we wszystkich obszarach zarządzania przedsiębiorstwem, szczególnie tam, gdzie liczy się szybkość podejmowania decyzji w odpowiedzi na zachodzące na rynku zmiany.

Bibliografia

1. Abramowicz W. (2003), *Filtrowanie informacji z Internetu do systemów informacyjnych zarządzania*, <http://repozytorium.uni.lodz.pl:8080/xmlui/bitstream/handle/11089/7059/7-18.pdf?sequence=1&isAllowed=y>.
2. Aleksander M., Decker J., Wehbe B. (2015), *Analizy Business Intelligence. Zaawansowane wykorzystanie Excela*, Helion, Gliwice.
3. Dudycz H (2010), *Visualization methods in Business Intelligence Systems – An Overview*, w: *Business Informatics. Data Mining and Business Intelligence*, J. Korczak (ed.), Research papers of Wrocław University of Economics, Wrocław.
4. GUS (2014), *Spółeczeństwo informacyjne w Polsce. Wyniki badań statystycznych z lat 2010–2014*, Urząd statystyczny w Szczecinie, <http://stat.gov.pl/obszary-tematyczne/nauka-i-technika-spoleczenstwo-informacyjne/spoleczenstwo-informacyjne/spoleczenstwo-informacyjne-w-polsce-wyniki-badan-statystycznych-z-lat-2010-2014,1,8.html>.
5. GUS (2015), <http://stat.gov.pl/statystyka-regionalna/jednostki-terytorialne/nomenklatura-nts/>.
6. Januszewski A. (2008), *Funkcjonalność informatycznych systemów zarządzania*, tom 1, PWN, Warszawa.
7. Masewicz M. (2009), *Zapewnianie jakości danych ładowanych do systemów analitycznych – omówienie możliwości narzędzi wbudowanych w Oracle Warehouse Builder 11g i Oracle Data Integrator 10g*, XV Konferencja PLOUG, Kościelisko, http://www.ploug.org.pl/konf_09/materialy/pdf/14_Zapewnianie_jakosci_danych_ladowanych.pdf.
8. PBSSP – Rozporządzenie Rady Ministrów z dnia 27 sierpnia 2014 r. w sprawie programu badań statystycznych statystyki publicznej na rok 2015 (Dz.U. poz. 1330).
9. Provost F., Fawcett T. (2013), *Data science and its relationship to big data and data-driven decision making*, „Big Data”, March 1(1), <http://online.liebertpub.com/doi/pdf/10.1089/big.2013.1508>.

10. Provost F., Fawcett T. (2015), *Analiza danych w biznesie. Sztuka podejmowania skutecznych decyzji*, Helion, Gliwice
11. Próchnicki W. (2014), *Zastosowanie tabel przestawnych w kontrolingu. Excel zaawansowany*, Tom 6, Wiedza i praktyka.
12. Surma J. (2009), *Business Intelligence: systemy wspomaganie decyzji biznesowych*, Wydawnictwo Naukowe PWN, Warszawa.
13. Walkenbach J. (2013), *Microsoft Excel 2013PL. Biblia*, Helion, Gliwice.
14. Walkenbach J., Aleksander M. (2014), *Analiza i prezentacja danych w Microsoft Excel*, Helion, Gliwice.
15. Winston W. L. (2014), *Microsoft Excel 2013: Analiza i modelowanie danych biznesowych*, APN Promise, Warszawa.
16. Ziara L. (2012), *Funkcjonalność kokpitów menedżerskich w systemach Business Intelligence. Przegląd wybranych rozwiązań*, w: *Systemy inteligencji biznesowej jako przedmiot badań ekonomicznych*, C. M. Olszak, E. Ziemia (red.), Zeszyty Naukowe Wydziałowe Uniwersytetu Ekonomicznego w Katowicach, Katowice 2012.
17. Ziuziański P. (2014), *Kokpit menedżerski jako efektywne narzędzie do wizualizacji danych w organizacji*, w: *Rola informatyki w naukach ekonomicznych i społecznych. Innowacje i implikacje interdyscyplinarne*, P. Zieliński (red.), Wydawnictwo Wyższej Szkoły Handlowej, Kielce.
18. Ziuziański P., Furmankiewicz M. (2015), *Rola kokpitu menedżerskiego w procesie podejmowania decyzji*, w: *Zeszyty Naukowe Politechniki Śląskiej, seria Organizacja i Zarządzanie*, nr 77.
19. <http://skuteczneraporty.pl/blog/wielokrotnosc-czyli-wykres-panelowy/>, 2015
20. <http://www.contextures.com/excelpanelchart.html>, 2015

Rozdział 6

Tworzenie modelu danych z wykorzystaniem Power Pivot

Dariusz Kralewski

Wstęp

Nowe technologie teleinformatyczne wymuszają zmiany w podejściu menedżera do zarządzania firmą. Systemy analityczne różnią się od dotychczasowych modeli wspomaganie decyzji technologią oraz sposobem wspomaganie decyzji. Zmiany po stronie technologii obejmują przede wszystkim hurtownie danych, zaawansowane techniki analityczne, techniki wizualizacji danych oraz systemy uczące się. Synergia wymienionych rozwiązań stwarza inteligentne środowisko do podejmowania decyzji w organizacji [Wrycza, 2010].

W każdej firmie wiedza jest kluczem do sukcesu. Dlatego też niezmiernie ważne jest, aby była ona pełna i opierała się na danych ze wszystkich obszarów działania przedsiębiorstwa. Stale powiększająca się liczba danych firmowych sprawia, że typowe informatyczne systemy transakcyjne nie wystarczają, by dokonać globalnej analizy, obejmującej wszystkie aspekty działalności przedsiębiorstwa. Dokładna analiza stanu firmy i jej otoczenia jest podstawą do oceny jej rentowności i prognozowania działań biznesowych. Aby osiągnąć zysk i przewagę konkurencyjną, szefowie firm muszą podejmować trafne decyzje. Decyzje te często muszą być podejmowane na podstawie niepełnych danych. Ponadto współczesne systemy transakcyjne nie są zoptymalizowane pod kątem przeprowadzenia analiz. Tworzenie wieloaspektowych statystyk może spowolnić działanie tychże systemów produkcyjnych [Januszewski, 2013].

Obecnie analitycy potrzebują narzędzia, które będzie zawierać wszystkie potrzebne funkcje niezbędne w analizie, a jednocześnie będzie potrafić prognozować na podstawie dostarczonych mu informacji. Co więcej, wyniki prac analityka powinny być raportowane w przejrzysty dla użytkowników sposób. Tego typu systemy określa się mianem „Inteligencji Biznesowej” (ang. *Business Intelligence*), a liczba ich zastosowań stale rośnie.

Środowiska informacyjne w firmie złożone są z wielu aplikacji i baz danych, tysiące dokumentów oraz dziesiątków tysięcy plików i folderów. Chociaż dane te wykorzystywane są przez aplikacje różnych producentów, wykonane w różnych technologiach, w różnym czasie, to są one najczęściej w jakiś sposób między sobą powiązane. Aplikacja analityczna musi pozwolić na określenie relacji między

danymi, które nie były dotychczas powiązane. Integracja danych jest procesem łączenia danych pochodzących z różnych źródeł. Umożliwia ona użytkownikowi ujednoczone postrzeganie tychże danych [Tsatalos i in., 1996]. Zanim dane poddane zostaną analizie, muszą ulec integracji, tzn. uwspólnieniu formy i treści pomiędzy różnorodnymi źródłami danych. Proces integracji danych jest jednym z najważniejszych procesów w całym cyklu życia danych. Proces ten obejmuje również takie operacje na danych, jak transformacje i czyszczenie danych, które opisano w rozdziale 2. Można na przykład połączyć dane o numerach telefonów kontrahentów z bilingiem połączeń telefonicznych, aby w ten sposób sprawdzić kontakty osób pracujących w firmie. Innym przykładem może być połączenie danych pochodzących z systemu CRM z systemem kadrowo-płacowym, aby przeanalizować adekwatność zarobków handlowców z ich wynikami. Łatwość integracji danych z różnych źródeł sprawia, że czas potrzebny na scalenie danych skracany jest do minimum.

Każdy system analityczny wyposażony jest w możliwość obsługi danych wielowymiarowych, czyli kostek OLAP. Dodatkowo wbudowane w system mechanizmy *data mining* pozwalają na wnioskowanie na podstawie danych historycznych. Dzięki takiemu podejściu analityk jest w stanie szybko przygotować analizę zdarzeń, które miały miejsce w przeszłości, oraz szybko wygenerować prognozę. Przygotowane wcześniej analizy powinny w automatyczny sposób znaleźć się w raporcie.

Aplikacje raportujące i analityczne komunikują się bezpośrednio z użytkownikiem i dostarczają mu przetworzonych danych w pożądanej formie. Można wyróżnić następujące grupy tych aplikacji:

- narzędzia raportujące,
- narzędzia OLAP,
- narzędzia do eksploracji danych,
- aplikacje analityczne,
- kokpity menedżerskie.

W obecnych czasach konieczne staje się poszukiwanie takich rozwiązań, które z jednej strony w zintegrowany sposób wspomagałyby wszystkie procesy związane z pozyskiwaniem, zarządzaniem i dystrybucją wiedzy wewnątrz organizacji, z drugiej zaś dawałyby możliwość tworzenia nowych sieci relacji [Surma, 2013].

Celem bieżącego rozdziału jest zobrazowanie tworzenia modelu danych z wykorzystaniem dodatku Power Pivot dla MS Excel. Poprawnie stworzony model pozwala na integrację danych pochodzących z różnych źródeł i jest podstawą analizy z wykorzystaniem takich narzędzi, jak: kokpit menedżerski, tabele i wykresy przestawne z wykorzystaniem Power Pivot, fragmentatory, pola obliczeniowe z wykorzystaniem języka DAX czy KPI. Zatem jest to niezwykle ważny etap przygotowania danych biznesowych do późniejszej analizy. Wszystkie opisa-

ne w niniejszym rozdziale mechanizmy mają zastosowanie w praktyce. Dla lepszego zilustrowania ich użyteczności poniżej przedstawiono studia przypadków.

6.1. Przypadek 1 – integracja kurów walut, danych giełdowych oraz danych sprzedaży

Firma X zamierza stworzyć kokpit menedżerski w celu analizy relacji notowań walut, danych giełdowych i sprzedaży. Informacje te są niezbędne do prawidłowego działania firmy y podejmowanych przez nią decyzji.

Wśród wielu rozwiązań umożliwiających stworzenie kokpitu menedżerskiego najłatwiejszym w implementacji dla rozważanej firmy był posiadany i wszechstronnie wykorzystywany MS Excel 2016. Power Pivot dla programu MS Excel jest narzędziem służącym do analizy danych, które zapewnia dużą moc obliczeniową bezpośrednio w użytkowanej aplikacji Microsoft Excel. Jest to przyjazny dla użytkownika sposób na przeprowadzenie analizy danych za pomocą znanych już możliwości arkusza kalkulacyjnego, takich jak tabele przestawne, widoki wykresów przestawnych i fragmentatory (ang. *slicers*) [Jelen, Collie, 2014]. Jest to szybki sposób na generowanie bogatych i interaktywnych narzędzi analizy, żeby uzyskać głębszy wgląd biznesowy i krótsze opóźnienia decyzyjne.

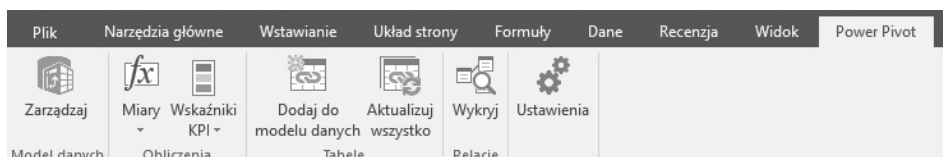
Korzystanie z rozszerzenia Power Pivot dla MS Excel ma wiele zalet dla użytkownika:

- korzystanie z dobrze znanych narzędzi i funkcji programu MS Excel,
- przetwarzanie ogromnych ilości danych w bardzo krótkim czasie,
- możliwość załadowania bardzo dużych zbiorów danych z prawie dowolnego źródła,
- możliwość użycia *Data Analysis Expressions* (DAX),
- pełne wykorzystanie procesorów wielordzeniowych i dużej ilości pamięci operacyjnej.

Power Pivot jest dodatkiem do programu MS Excel, który umożliwia wykonywanie zaawansowanych analiz danych oraz tworzenie złożonych modeli danych. Dodatek Power Pivot umożliwia pracę z dużymi ilościami danych z różnych źródeł, błyskawiczne analizowanie informacji oraz łatwe udostępnianie wyników analiz. Zarówno w programie MS Excel, jak i w dodatku Power Pivot można utworzyć model danych, czyli kolekcję tabel połączonych relacjami. Model danych widoczny w skoroszytcie programu MS Excel jest taki sam, jak model danych widoczny w oknie dodatku Power Pivot. Wszystkie dane zaimportowane do programu MS Excel są dostępne w dodatku Power Pivot i odwrotnie.

Dodatek Power Pivot w programie Microsoft Excel 2016 umożliwia wykonywanie zaawansowanych analiz. Jest on wbudowany w program MS Excel 2016

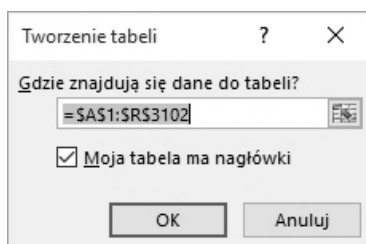
i charakteryzuje się oddzielną zakładką Power Pivot. W odróżnieniu od wcześniejszych wersji jest on już standardowo włączony. Gdyby jednak nie było odpowiedniej wstążki (rysunek 6.1), należy sprawdzić, czy włączony jest dodatek Power Pivot (wstążka *Developer* > *Dodatki COM lub Plik* > *Opcje* > *Dodatki* i > *Dodatki COM*). Zaznaczone pole *Microsoft Power Pivot for Excel* świadczy o poprawnym załadowaniu modułu.



Rysunek 6.1. Widok wstążki Power Pivot

Źródło: Opracowanie własne.

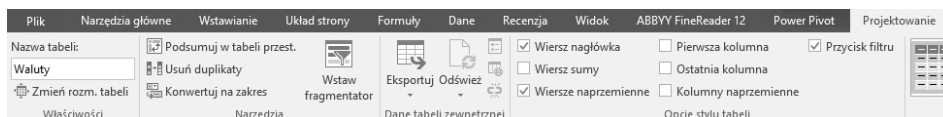
Zanim przystąpi się do analizy danych za pomocą Power Pivot, należy oznaczyć i nazwać poszczególne tabele. Tworzenie tabel pozwala na organizowanie i późniejsze analizowanie powiązanych danych. Tabele ponadto ułatwiają sortowanie, filtrowanie i formatowanie danych w arkuszu. Wybór narzędzia *Tabela* skutkuje pojawieniem się okna dialogowego, w którym należy zaznaczyć/sprawdzić odpowiedni zakres danych i oznaczyć opcję wskazującą, że zaznaczone dane posiadają nagłówki (rysunek 6.2).



Rysunek 6.2. Tworzenie tabeli w programie Microsoft Excel

Źródło: Opracowanie własne.

Niezmiernie ważnym kolejnym krokiem jest nadanie nazwy nowo utworzonej tabeli (rysunek 6.3). Nazwy te będą pojawiały się w relacjach, będą także wykorzystywane w formułach obliczeniowych.

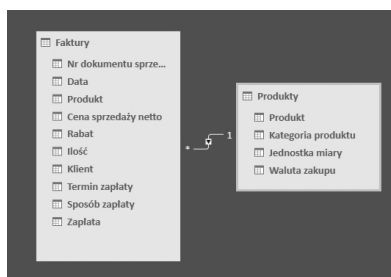


Rysunek 6.3. Nazwanie nowo utworzonej tabeli

Źródło: Opracowanie własne.

Firma, mając utworzone tabele, powinna powiązać odpowiednie dane. Podobnie jak w relacyjnej bazie danych, zbiór danych to opisany i zorganizowany zbiór tabel połączonych relacjami, czyli związkami między sobą. Ten sposób przechowywania informacji pozwala na uniknięcie redundancji (powtarzania się danych) oraz przeprowadzanie analiz na podstawie wielu tabel. Każda tabela składa się z rekordów (pojedynczych wierszy) [Jelen, Collie, 2015]. Poszczególne rekordy składają się z pól (komórek) przechowujących atomową daną. Aby istniała możliwość utworzenia z tabel relacyjnego modelu danych, przynajmniej w jednej z nich musi występować klucz główny (zwany też podstawowym) – kolumna zawierająca wartości służące do identyfikacji poszczególnych rekordów tabeli [Todman, 2003]. Wartości w kluczu podstawowym muszą być zatem unikalne. Zazwyczaj funkcję takiego klucza spełnia numer porządkowy. Wszelkie tabele muszą być powiązane wspólnymi danymi z odpowiednich kolumn.

Fragment modelu danych łączącego informacje o fakturach sprzedaży (tabela *Faktury*) z informacjami o produktach (tabela *Produkty*) utworzonego w Power Pivot prezentuje rysunek 6.4.



Data	Produkt	Cena sprz	Rabat	Ilość		Produkt	Kategoria	Jednostka
07.01.2014	Jabłka świeże	2,00 zł	5%	115		Jabłka świeże	Owoce	kg
07.01.2014	Gruszki świeże	4,00 zł	10%	184		Gruszki świeże	Owoce	kg
07.01.2014	Banany	3,40 zł	0%	20		Banany	Owoce	kg
07.01.2014	Marchew	1,50 zł	5%	109		Marchew	Warzywa	kg
07.01.2014	Pomidor	5,00 zł	0%	57		Pomidor	Warzywa	kg
07.01.2014	Ziemniaki	3,00 zł	0%	65		Ziemniaki	Warzywa	kg
08.01.2014	Pomarańcza	4,50 zł	0%	65		Pomarańcza	Owoce	kg
08.01.2014	Nektarynki	3,00 zł	10%	196		Nektarynki	Owoce	szt
08.01.2014	Brzoskwinie	2,40 zł	0%	24		Brzoskwinie	Owoce	szt
08.01.2014	Winogrona	8,10 zł	5%	107		Winogrona	Owoce	kg
08.01.2014	Truskawki	12,00 zł	5%	118		Truskawki	Owoce	kg
08.01.2014	Jabłka świeże	2,00 zł	5%	141				
09.01.2014	Gruszki świeże	4,00 zł	5%	131				
09.01.2014	Banany	3,40 zł	0%	57				
09.01.2014	Marchew	1,30 zł	0%	49				
09.01.2014	Pomidor	5,00 zł	5%	127				
09.01.2014	Ziemniaki	3,00 zł	10%	176				
09.01.2014	Pomarańcza	4,50 zł	0%	71				
10.01.2014	Nektarynki	3,00 zł	0%	85				
10.01.2014	Brzoskwinie	2,50 zł	0%	31				
10.01.2014	Winogrona	8,40 zł	5%	146				
10.01.2014	Truskawki	12,00 zł	8%	175				

Rysunek 6.4. Połączenie tabel za pomocą klucza

Źródło: Opracowanie własne.

W powyższym przykładzie występuje tzw. klucz prosty – identyfikacja rekordu następuje za pomocą pojedynczego pola. Istnieją także klucze złożone. W ich przypadku identyfikacja wiersza dokonuje się za pomocą więcej niż jednej kolumny. W Power Pivot nie ma możliwości zastosowania klucza złożonego [Alexander, Decker, 2014].

Model opisują poszczególne połączone tabele. Relacje tworzone są poprzez przeciąganie i przez to łączenie odpowiednich nazw kolumn. Przy tej czynności nie ma znaczenia kierunek przeciągania (taki sam efekt otrzyma się, przeciągając nazwę kolumny z tabeli A do B, jak i z B do A) [Czapiewski, 2014a].

Relację ustanawia się pomiędzy dwoma tabelami, na podstawie wartości klucza podstawowego w jednej tabeli i kolumny w drugiej tabeli zawierającej wartości klucza podstawowego z tabeli pierwszej. Wyróżnia się trzy rodzaje relacji [Czapiewski, 2014b]:

1. Jeden do jednego – w tego typu relacji jednemu rekordowi z tabeli A odpowiada tylko jeden wiersz z tabeli B. Rodzaj ten występuje stosunkowo rzadko, ponieważ wszystkie informacje przechowywane w ten sposób można zamieścić w jednej tabeli.
2. Jeden do wielu – jednemu rekordowi z tabeli A odpowiada wiele rekordów z tabeli B. Jest to najpowszechniejszy typ relacji.
3. Wiele do wielu – rekord w tabeli A może mieć dopasowanych wiele wierszy z tabeli B oraz odwrotnie – rekord z tabeli B może mieć dopasowanych wiele wierszy z tabeli A. Taki typ jest możliwy do zdefiniowania tylko poprzez dodanie do modelu trzeciej tabeli (zwanej tabelą łącząca), w której będą znajdowały się wartości kluczy podstawowych tabel A oraz B.

Dzięki dodatkowi Power Pivot można stworzyć relacje w sposób prosty, intuicyjny i bardzo poglądowy [Ferrari, Russo, 2014]. W pierwszej kolejności należy dodać tabele do modelu – wstążka *Power Pivot* > *Dodaj do modelu danych* (rysunek 6.5).



Rysunek 6.5. Wstążka Power Pivot

Źródło: Opracowanie własne.

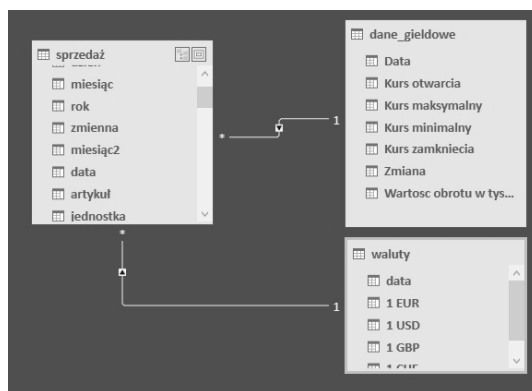
W dodatku Power Pivot tabele można przeglądać w dwóch widokach (wstążka *Narzędzia główne*): w *Widoku danych* lub w *Widoku diagramu*. W pierwszym przypadku prezentowana jest pojedyncza tabela z danymi oraz ewentualnymi, jeśli zostały zdefiniowane i włączone (*Obszar obliczeń*), polami obliczeniowymi (miarami).

W drugim przypadku widoczne są dodane do modelu tabele wraz z relacjami pomiędzy nimi (np. rysunek 6.6).

W przypadku Firmy X należało zintegrować tabele wygenerowane w rozdziale 1 i odpowiednio sformatowane w rozdziale 2, z danymi dotyczącymi:

- notowań kursów walut,
- notowań giełdowych,
- sprzedaży.

W tym celu w dodatku Power Pivot pobrano odpowiednie tabele (wstążka *Narzędzia główne* > *Pobieranie danych zewnętrznych* > *Z innych źródeł* > *Plik programu Excel*) i nawiązano relacje pomiędzy tabelami. Należy zauważyć, że kluczem łączącym wszystkie trzy tabele było pole *Data*. Stworzony model danych zaprezentowany został na rysunku 6.6 (*r6_sprzedaz_gielda_waluty_model.xlsx*).



Rysunek 6.6. Model danych – integracja indeksów giełdowych, kursów walut i danych dotyczących sprzedaży

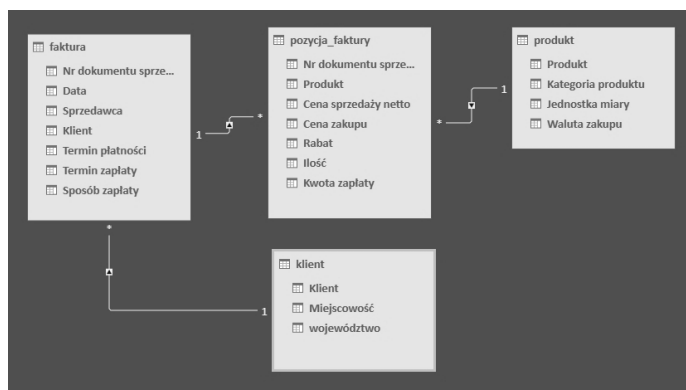
Źródło: Opracowanie własne.

Proces tworzenia kokpitu menedżerskiego z wykorzystaniem modelu danych zostanie szczegółowo opisany w rozdziale 7.

6.2. Przypadek 2 – integracja danych dotyczących sprzedaży

Firma X w codziennej swojej pracy wykorzystuje arkusz kalkulacyjny MS Excel do analizy sprzedaży na podstawie danych pochodzących z faktur. Informacje pozyskiwane są z różnych systemów transakcyjnych i umieszczane w arkuszu. Plik arkusza kalkulacyjnego staje się niebezpiecznie duży, a ogromna liczba wierszy powoduje brak czytelności. Dodatkowo pracownicy zauważyli, iż część danych jest niepotrzebnie powielana, co powoduje niepotrzebny wysiłek i możliwość pomyłki. Zaczęto zastanawiać się, czy nie można by było jakoś uprościć,

uporządkować (znormalizować) wprowadzanych danych. Z tego też powodu arkusz z fakturami (plik *r2_faktury_sprzedazy_o.xlsx*) podzielono na kilka mniejszych arkuszy: *Faktury sprzedaży*, *Pozycje faktur*, *Klienci*, *Produkty* (por. rozdział 2, przypadek 3). Arkusze te służą jako dane źródłowe do przeprowadzania analiz sprzedaży za pomocą tabel i wykresów przestawnych (tworzenie tabel przestawnych na podstawie kilku arkuszy omówione zostanie szerzej w rozdziale 7). Należało je jednak wcześniej powiązać, czyli utworzyć odpowiedni model danych w dodatku Power Pivot. Efekt złączenia został zaprezentowany na rysunku 6.7 (plik *r6_faktury_sprzedazy_model.xlsx*).



Rysunek 6.7. Relacje pomiędzy tabelami z danymi na temat sprzedaży

Źródło: Opracowanie własne.

Zanim nastąpiło połączenie, dane w poszczególnych arkuszach MS Excel oznaczono jako tabele oraz je odpowiednio nazwano (wstążka *Projektowanie > Nazwa tabeli*). Następnie dane skopiowano do Power Pivot, gdzie zdefiniowano połączenia pomiędzy tabelami. Należy zauważyć, że kluczem łączącym arkusze *Faktura* – *Pozycja faktury* był *Nr dokumentu sprzedaży*, kluczem łączącym tabele *Klient* – *Faktura* było pole *Klient*, a kluczem łączącym *Produkt* – *Pozycja faktury* było pole *Produkt*. Dzięki zdefiniowaniu właściwych relacji pomiędzy tabelami dane są zintegrowane i spójne, co usprawnia ich analizę oraz niweluje ryzyko popełniania błędów, szczególnie podczas ewentualnych korekt danych.

6.3. Przypadek 3 – wykorzystanie języka DAX w analizie danych na temat sprzedaży

Firma X wykorzystuje arkusz z poprzedniego studium przypadku do analizy sprzedaży. Podczas realizacji bieżących czynności analitycznych pojawiła się jednak potrzeba stworzenia dwóch dodatkowych kolumn obliczeniowych:

- *Zapłata* – która to byłaby odpowiednią operacją na kolumnach: *Cena sprzedaży netto*, *Podatek VAT*, *Ilość*, *Waluta zakupu*;
- *Ilość z jednostką* – która byłaby odpowiednią operacją na kolumnach: *Ilość* oraz *Jednostka miary*.

Używając kolumn obliczeniowych, można dodawać nowe dane do tabeli w modelu danych dodatku Power Pivot. Jednak zamiast wklejać lub importować wartości do kolumny, należy utworzyć formułę, używając języka DAX (*Data Analysis Expressions*), która definiuje wartości w kolumnie. Formuły w kolumnach obliczeniowych są bardzo podobne do formuł tworzonych w programie MS Excel. Jednak w przeciwieństwie do MS Excel nie można tworzyć różnych formuł dla poszczególnych wierszy tabeli. Formuła języka DAX jest automatycznie stosowana do całej kolumny. Gdy kolumna zawiera formułę, wartość jest obliczana dla każdego wiersza zaraz po wprowadzeniu formuły. Wartości w kolumnie są obliczane, gdy jest to konieczne, na przykład w przypadku odświeżenia danych źródłowych. Kolumny obliczeniowe można również tworzyć na podstawie pól obliczeniowych i innych kolumn obliczeniowych. Można na przykład utworzyć jedną kolumnę obliczeniową w celu wyodrębnienia liczby z ciągu tekstowego, a następnie użyć tej liczby w innej kolumnie obliczeniowej.

Domyślnie nowe kolumny obliczeniowe są dodawane z prawej strony innych kolumn, a kolumnie automatycznie przypisywana jest nazwa. Można jednak zmienić rozmieszczenie oraz nazwy utworzonych kolumn. Istnieją pewne ograniczenia dotyczące zmieniania nazw kolumn obliczeniowych:

1. Każda nazwa kolumny musi być unikatowa w tabeli.
2. Należy unikać nadawania nazw takich samych jak nazwy pól obliczeniowych, które są używane w tym samym skoroszycie. Mimo że jest możliwe, aby pole obliczeniowe i kolumna obliczeniowa miały taką samą nazwę, użycie nie-unikatowych nazw może spowodować błędy w obliczeniach. Aby uniknąć przypadkowego wywołania pola obliczeniowego, podczas odwoływania się do kolumny należy zawsze używać w pełni kwalifikowanego odwołania do kolumny.
3. W przypadku zmiany nazwy kolumny obliczeniowej muszą zostać zaktualizowane wszelkie formuły zależne od tej kolumny. Jeśli nie jest włączony tryb aktualizacji ręcznych, wyniki formuł są aktualizowane automatycznie. Jednak ta operacja może zająć trochę czasu.

Formuła dla kolumny obliczeniowej może wymagać większej ilości zasobów niż formuła dla pola obliczeniowego. Jedną z przyczyn tego faktu jest to, że wynik kolumny obliczeniowej jest zawsze obliczany dla każdego wiersza w tabeli, podczas gdy pole obliczeniowe jest obliczane tylko dla komórek używanych w tabeli przestawnej lub na wykresie przestawnym. Na przykład tabela zawierająca milion wierszy zawsze będzie miała kolumnę obliczeniową zawierającą milion wyników,

co ma odpowiedni wpływ na wydajność. Jednak ogólnie tabela przestawna filtruje dane, stosując nagłówki wierszy i kolumn, więc pole obliczeniowe jest obliczane tylko dla podzbioru danych w każdej komórce tabeli przestawnej.

Formuła zwykle zawiera zależności od obiektów, do których się odwołuje, takich jak inne kolumny czy wyrażenia obliczające wartości. Domyślnie jest włączone odświeżanie automatyczne i dlatego wszystkie takie zależności mogą mieć wpływ na wydajność, gdy wartości są aktualizowane, a formuły odświeżane. Aby uniknąć problemów z wydajnością związanych z tworzeniem kolumn obliczeniowych, należy postępować zgodnie z następującymi wskazówkami:

1. Zamiast tworzyć jedną formułę zawierającą wiele złożonych zależności, należy tworzyć formuły krokowo z zapisywaniem ich wyników w kolumnach, co umożliwi sprawdzanie poprawności wyników i ocenę wydajności.
2. Modyfikacja danych często wymaga ponownego przeprowadzenia obliczeń w kolumnach obliczeniowych. Można temu zapobiec, ustawiając ręczny tryb ponownego obliczania, jednak jeśli dowolne wartości w kolumnie obliczeniowej będą niepoprawne, kolumna zostanie wyszarzona do czasu odświeżenia i ponownego obliczenia danych przez użytkownika.
3. Zmiana lub usunięcie relacji między tabelami może spowodować, że formuły używające kolumn znajdujących się w tych tabelach staną się nieprawidłowe.
4. Utworzenie formuły zawierającej zależność cykliczną lub odwołującą się do samej siebie powoduje wystąpienie błędu.

Najczęściej stosowaną funkcją jest funkcja RELATED. Funkcja ta zwraca powiązaną wartość z innej tabeli. Wartością zwracaną jest pojedyncza wartość powiązana z bieżącym wierszem. Funkcja RELATED wymaga, aby istniała relacja między bieżącą tabelą a tabelą z informacjami powiązanymi. Należy określić kolumnę zawierającą żądane dane, a funkcja śledzi istniejącą relację wiele-do-jednego w celu pobrania wartości z określonej kolumny w tabeli powiązanej. Jeśli relacja nie istnieje, trzeba ją utworzyć. Gdy funkcja RELATED przeprowadza wyszukiwanie, analizuje wszystkie wartości w określonej tabeli, niezależnie od wszelkich zastosowanych filtrów. Funkcja RELATED wymaga kontekstu wiersza, dlatego można jej używać w wyrażeniu kolumny obliczeniowej, w którym kontekst wiersza jest jednoznaczny.

6.4. Przypadek 4 – integracja danych dotyczących importu i eksportu owoców

W rozdziale 1 Firma X pozyskała (przypadek 1), a w rozdziale 2 firma przygotowała do analizy (przypadek 1) dane dotyczące importu i eksportu owoców w la-

tach 2013 oraz 2014 (plik *r2_import_i_eksport_o.xlsx*). Na podstawie tychże danych firma pragnie sprawdzić korelację między importem i eksportem poszczególnych owoców do poszczególnych krajów, sprawdzić trend oraz wyznaczyć prognozę na rok 2015. Aby tego dokonać, należało połączyć ze sobą dane zawarte w poszczególnych arkuszach/tabelach. Polami łączącymi tabele były kody CN owoców i nazwy poszczególnych państw. Aby móc dokonać połączenia na podstawie tychże pól, stworzono dwa oddzielne arkusze, w których znajdowały się unikalne kody owoców i nazwy państw. Końcowy efekt połączenia widoczny jest na rysunku 6.8 (plik *r6_import_i_eksport_model.xlsx*).



Rysunek 6.8. Relacje pomiędzy tabelami zawierającymi dane dotyczące importu i eksportu owoców

Źródło: Opracowanie własne.

Tabel import2013, eksport2013, import2014, eksport2014 firma nie była w stanie bezpośrednio połączyć. Firma pragnie analizować zmiany importu i eksportu określonych owoców w poszczególnych państwach. Zatem polami, które łączyły tabele, były *KodCN* i *Kraj*. Jednak próba połączenia bezpośrednio tychże pól skutkowała błędem, gdyż Power Pivot nie był w stanie połączyć z sobą wielu tych samych pozycji. W tej sytuacji firma postanowiła skopiować nazwy krajów i kody owoców do oddzielnych, odpowiednio nazwanych tabel, gdzie usunięto duplikaty danych. Sposób ten pozwolił na poprawne połączenie tabel.

6.5. Możliwe zastosowania MS Excel w biznesie w formie zadań

W celu zobrazowania tworzenia modelu danych do analizy z wykorzystaniem narzędzia MS Excel można posłużyć się poniższymi zadaniami:

1. Firma X zajmuje się sprzedażą owoców i warzyw. Firma posiada dane na temat swoich transakcji sprzedaży (*r2_faktury_sprzedazy_o.xlsx*). Główny Urząd

- Statystyczny co miesiąc publikuje dane na temat cen produktów żywnościowych. Firma pragnie sprawdzić, jakie są odchylenia cen sprzedaży od średnich krajowych.
2. Poszczególne owoce i warzywa posiadają różne stawki podatku VAT. Firma X musi utworzyć i dołączyć tabelę z podatkami do tabel zawierających dane dotyczące sprzedaży (rozważanych w przypadku 2), tak aby móc poprawnie wyliczyć cenę brutto sprzedaży na poszczególnych fakturach.
 3. W ramach przypadku 4 Firma X poddała analizie import i eksport owoców w latach 2013, 2014. Należy dołączyć dane z roku 2015 i wcześniejszych lat, tak aby prognozy były trafniejsze.
 4. Firma X, w ramach przypadku 4, połączyła tabele z danymi dotyczącymi importu i eksportu owoców w latach 2013 i 2014. W tabelach kraje i kody należy stworzyć odpowiednie kolumny obliczeniowe wskazujące na zsumowane wielkości importu i eksportu w poszczególnych krajach i poszczególnych owoców.

Zakończenie

Niniejszy rozdział pokazuje, iż MS Excel jest narzędziem pozwalającym na skuteczną analitykę biznesową. Power Pivot dla programu MS Excel pozwala w przyjazny i łatwy sposób dokonać analizy danych za pomocą znanych funkcjonalności arkusza kalkulacyjnego, takich jak tabele przestawne, widoki wykresów przestawnych czy fragmentatory. Program ten pozwala na głębszy wgląd biznesowy na podstawie dostarczonych danych, co skutkuje trafniejszymi decyzjami i krótszymi opóźnieniami decyzyjnymi. Umiejętne wykorzystanie prezentowanych mechanizmów jest niezbędną umiejętnością, którą powinien posiadać każdy analityk biznesowy.

Powyższe rezultaty jednoznacznie wskazują, że MS Excel jest bardzo użytecznym narzędziem dla zaawansowanej analizy danych. Ważne jest, by analizę danych postrzegać jako całościowy, wieloetapowy proces [Winston, 2014]. W pierwszym kroku należy zadbać o pozyskanie wysokiej jakości danych i przeniesienie ich do MS Excel (opisano ten etap w rozdziale 1). Dane te najczęściej nie nadają się od razu do analizy. Należy zatem je oczyścić i odpowiednio sformatować (opisano ten etap w rozdziale 2). Dopiero poprawnie sformatowane dane są podstawą do dalszych działań analitycznych. Jeśli dane te znajdują w różnych skoroszytach, bądź są pozyskiwane z różnych źródeł, należy je połączyć z sobą. Poprawnie połączone tabele są źródłem do dalszej analizy. Kolejny rozdział jest naturalnym rozwinięciem teorii, zastosowań i przypadków podejmowanych w ni-

niejszym rozdziale. Zintegrowane z różnych źródeł dane będą podstawą analiz realizowanych między innymi za pośrednictwem kokpitu menedżerskiego, którego proces tworzenia zaprezentowano w rozdziale 7.

Bibliografia

1. Alexander M., Decker J. (2014), *Microsoft Business Intelligence Tools for Excel Analysts*, Wiley.
2. Czapiewski B. (2014a), *Czym jest relacyjna baza danych? Jak tworzyć relacje w PowerPivot?*, <http://powerview.pl/czym-jest-relacyjna-baza-danych-jak-tworzyc-relacje-w-power-pivot/>.
3. Czapiewski B. (2014b), *Co to jest KPI w PowerPivot i jak go wstawić?*, <http://powerview.pl/tag/powerpivot/feed/>.
4. Ferrari A., Russo M. (2014), *Microsoft Excel 2013. Budowanie modeli danych przy użyciu PowerPivot*, Promise, Warszawa.
5. Januszewski A. (2013), *Funkcjonalność informatycznych systemów zarządzania*, t. 2, *Systemy Business Intelligence*, PWN, Warszawa.
6. Jelen B., Collie R. (2015), *Power Pivot dla Excela. Zaawansowane możliwości*, Helion, Gliwice.
7. Jelen B., Collie R. (2014), *PowerPivot Alchemy: Patterns and Techniques for Excel*, Holy Macro! Books.
8. Surma J. (2013), *Business Intelligence, Systemy wspomaganie decyzji biznesowych*, PWN, Warszawa.
9. Todman Ch. (2003), *Projektowanie hurtowni danych. Zarządzanie kontaktami z klientami (CRM)*, Wydawnictwa Naukowo-Techniczne, Warszawa.
10. Tsatalos O.G., Solomon M.H., Ioannidis Y.E. (1996), *The GMAP: A versatile tool for physical data independence*, „Very Large Database Journal”, Vol. 5, No. 2.
11. Winston W. (2014), *Microsoft Excel 2013 Data Analysis and Business Modeling*, Microsoft Press.
12. Wrycza S. (red.) (2010), *Informatyka ekonomiczna. Podręcznik akademicki*, PWE, Warszawa.

Rozdział 7

Podstawy tworzenia kokpitu menedżerskiego z wykorzystaniem Power Pivot

Jacek Maślankowski

Wstęp

W warunkach panujących we współczesnym biznesie kluczowy wydaje się szybki dostęp do informacji. Decydenci nie dysponują wystarczającą ilością czasu na czytanie długich i rozbudowanych raportów. Z tego powodu coraz większego znaczenia nabierają kokpity menedżerskie, zwane również pulpitami menedżerskimi, których głównym celem jest przekazanie w możliwie przystępnej formie, umożliwiającej szybko i trafną diagnozę, zmian zachodzących w biznesie.

Kokpit menedżerski jest jednym z narzędzi zarządzania, wspomagającym podejmowanie decyzji poprzez dostarczanie kompletnych danych na temat danego zjawiska [Guni, 2014]. Jest to zestaw mierników prezentowanych zazwyczaj w formie graficznej, prezentujący najważniejsze informacje z wybranego obszaru działalności organizacji [Bremser, Wagner, 2013]. Może być wykorzystywany np. do pomiaru wydajności produkcji w przedsiębiorstwie. Takie mierniki są często nazywane kluczowymi wskaźnikami wydajności (ang. *key performance indicators* – KPI) i mają zastosowanie w wielu dziedzinach działalności przedsiębiorstwa, np. w monitorowaniu zachowań klienta [Lautman, Pauwels, 2013].

Kokpity menedżerskie nie są nową koncepcją. Od dawna dostępne są aplikacje, zwłaszcza w przemyśle, pozwalające na monitorowanie wybranych wskaźników, takich jak np. bieżące zużycie prądu, temperatura napędu, pozostający poziom energii [Homocianu, Airinei, 2015]. Kokpit menedżerski powszechnie jest wykorzystywany do tworzenia podsumowania aktualnego wyniku finansowego przedsiębiorstwa [Manzetti, Mehta, 2015]. Równie dobrze może być wykorzystywany jako narzędzie zarządzania projektami z wykorzystaniem zwinnych metodyk zarządzania [Sangeeta, Malarvizhi, 2015]. Najczęściej jednak kokpit menedżerski utożsamiany jest z pojęciem analityki biznesowej [Bradea i in., 2014]. Dlatego też narzędzia te wchodzi w skład systemów informatycznych klasy Business Intelligence, których celem jest dostarczanie decydującym niezbędnych informacji do wspomagania podejmowania decyzji [Hsinchun i in., 2012]. Systemy te znajdują powszechne zastosowanie w biznesie [Wrycza, 2010].

Trzy podstawowe cechy, jakie musi spełniać kokpit menedżerski, to: relewantność (spełnianie celu, w jakim został utworzony), pilność (aktualność i szybkość

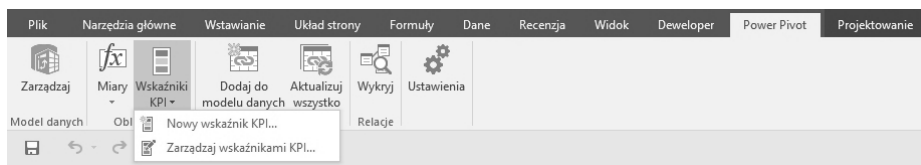
dostarczenia informacji) oraz efektywność (przejrzysty instrument prezentacji najważniejszych danych). Niezwykle ważne jest zatem odpowiednie dobranie zestawu danych, jakie mają być prezentowane w kokpicie [Guni, 2014]. Równie istotne przy doborze kokpitu menedżerskiego są trzy aspekty techniczne: intuicyjna grafika, wykorzystane narzędzia analityczne oraz łatwość kustomizacji, czyli wprowadzania zmian w celu dostosowania kokpitu do indywidualnych wymagań decydenta [Shadpour, Kilcoyne, 2015].

Celem niniejszego rozdziału jest zaprezentowanie możliwości wsparcia decydenta poprzez nowoczesne formy prezentacji danych biznesowych, jakimi są kokpity menedżerskie. Dokonano klasyfikacji kokpitów oraz zwrócono uwagę na najczęściej popełniane podczas ich tworzenia błędy. Przygotowano również rozwiązania, które umożliwiają zastosowanie tego rodzaju kokpitów menedżerskich w niemal każdym prowadzonym biznesie. Poza danymi biznesowymi w opisywanych studiach przypadków wykorzystano również źródła danych, które są powszechnie dostępne w sieci Internet.

7.1. Przypadek 1 – interaktywny kokpit do analizy notowań giełdowych i kursów walut

W niniejszym studium przypadku rozważono przypadek Firmy X, która prowadzi sprzedaż owoców i warzyw. Ponieważ produkty miały być eksportowane, jak również importowane w ramach granic Unii Europejskiej, kluczowe wydawały się wskaźniki dotyczące siły nabywczej złotówki. Firmie tej zależało również na informacji o stanie polskiej giełdy (wartości indeksów giełdowych) oraz przede wszystkim na aktualnym kursie walutowym. W tym celu zaproponowano utworzenie kokpitu menedżerskiego, który miał zawierać najbardziej aktualne informacje w tym zakresie. Dodatkowo miała również istnieć możliwość przeglądania archiwum danych.

Istnieje wiele narzędzi do tworzenia kokpitów menedżerskich, które najczęściej utożsamiane są z rozbudowanymi systemami typu Business Intelligence. W opisywanym przypadku zdecydowano się jednak na zastosowanie narzędzia MS Excel z dodatkiem Power Pivot. Przykładowe opcje tego dodatku zostały zaprezentowane na rysunku 7.1.



Rysunek 7.1. Narzędzia Power Pivot w arkuszu kalkulacyjnym MS Excel

Źródło: Opracowanie własne.

Na rysunku 7.1. można zauważyć, że jedną z funkcji tego dodatku jest umożliwienie definiowania tzw. kluczowych wskaźników wydajności – KPI, opisywanych we wstępie do niniejszego rozdziału. Definiowanie tych wskaźników jest możliwe po wcześniejszym zdefiniowaniu pól obliczeniowych. W tym przypadku pola obliczeniowe będą zawierały formuły, według których wyliczane będą wskaźniki KPI.

Menedżerowie firmy nie byli zdecydowani co do formy prezentacji danych. Zasadniczo zaproponowane dla nich kokpity menedżerskie można było przyporządkować do jednego z trzech poniższych poziomów:

- 1) statyczny kokpit z danymi,
- 2) wyświetlanie danych z częściową interaktywnością,
- 3) zintegrowana kontrola i charakter analityczny kokpitu.

W niniejszym studium przypadku zaprezentowano kokpit z częściową interaktywnością. Zarząd zdecydowanie poparł propozycję utworzenia kokpitu prezentującego dane w formie tabelarycznej, jak również wykresów.

Poniżej opisano proponowane etapy tworzenia kokpitu menedżerskiego. Zakłada się, że dane zostały już oczyszczone i przygotowane, zgodnie z regułami opisanymi w rozdziale 2 niniejszej książki. Zatem proponowane etapy tworzenia kokpitu menedżerskiego będą następujące:

- 1) wczytanie danych (polecenie *Pobieranie danych zewnętrznych* lub *Dodaj do modelu danych*),
- 2) znalezienie wspólnej kolumny (klucza), według którego można połączyć zbiory – dotyczy to tylko wielu zbiorów danych,
- 3) połączenie zbiorów danych,
- 4) utworzenie zestawu tabel i wykresów,
- 5) przygotowanie wskaźników KPI.

Na rysunku 7.2. zaprezentowano tabelę wczytaną do narzędzia Power Pivot z poziomu MS Excel poprzez opcję *Dodaj do modelu danych* (wstążka *Power Pivot*).

The screenshot shows the PowerPivot for Excel interface. The ribbon includes 'Narzędzia główne', 'Projekt', 'Zaawansowane', and 'Połączona tabela'. The 'Narzędzia główne' group contains options like 'Wklej i dołącz', 'Wklej i zamień', 'Kopiuuj', and 'Schowek'. The 'Zaawansowane' group includes 'Z bazy danych', 'Z usługi danych', 'Z innych źródeł', 'Istniejące połączenia', 'Odśwież', and 'Tabela przestawna'. The data table below has columns for 'data', '1 EUR', '1 USD', '1 GBP', '1 CHF', and 'Dodaj kolumnę'. The data rows show dates from 2015-05-15 to 2015-06-03 and corresponding exchange rates.

	data	1 EUR	1 USD	1 GBP	1 CHF	Dodaj kolumnę
	2015-06-03	4,1307	3,7108	5,6682	3,969	
	2015-06-02	4,1374	3,7676	5,744	4,0019	
	2015-06-01	4,126	3,785	5,7658	3,9984	
	2015-05-29	4,1301	3,7671	5,759	3,991	
	2015-05-28	4,1419	3,7858	5,8059	4,0014	
	2015-05-27	4,1405	3,7906	5,8407	4,0014	
	2015-05-26	4,1279	3,7898	5,8386	3,9864	
	2015-05-25	4,1149	3,75	5,8063	3,9757	
	2015-05-22	4,098	3,669	5,7447	3,9355	
	2015-05-21	4,0836	3,6605	5,733	3,926	
	2015-05-20	4,0606	3,6538	5,6685	3,8977	
	2015-05-19	4,0465	3,6152	5,6194	3,8787	
	2015-05-18	4,0468	3,555	5,5672	3,8642	
	2015-05-15	4,0587	3,5719	5,6269	3,8955	

Rysunek 7.2. Zbiór danych wczytany do narzędzia Power Pivot

Źródło: Opracowanie własne.

Z perspektywy użytkownika kokpitu menedżerskiego niezwykle istotny jest właściwy opis danych. Trzeba pamiętać, aby dane źródłowe miały w nagłówku nazwy kolumn. Bardzo to ułatwia późniejszą nawigację analitykom danych, którzy będą pracowali na zbiorze.

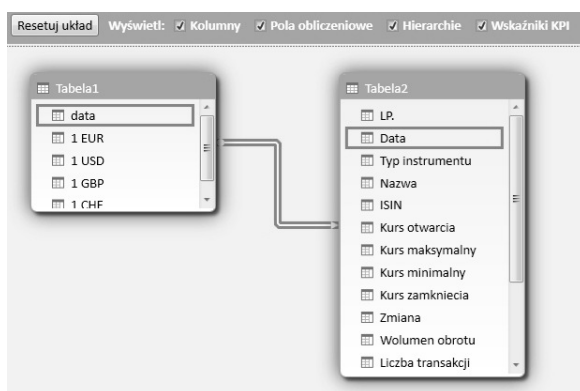
Do największych wyzwań związanych z realizacją opisywanego w studium przypadku kokpitu menedżerskiego należało ujednoczenie i połączenie danych nt. kursu walut oraz notowań giełdowych, a następnie przygotowanie ich do prezentacji w postaci jednego zbioru danych. Zatem jednym z pierwszych zadań, jakie jest zwykle realizowane przy tworzeniu kokpitu menedżerskiego, jest znalezienie części wspólnych zbioru danych. W niniejszym przypadku wykorzystano zbiór danych z walutami (plik *r1_kursy_walut.xlsx*) oraz z danymi giełdowymi (plik *r1_dane_gieldowe.xlsx*). Jako klucz połączeń wielu zbiorów danych wykorzystuje się najczęściej datę lub uznane klasyfikacje standardowe, np. kody krajów według normy ISO-3166. Można też tworzyć swoje własne klasyfikacje, przypisując kody liczbowe do odpowiednich wartości (np. 1 – Polska, 2 – Niemcy itd.).

Analizując możliwość połączenia tego rodzaju danych, zdecydowano się na powiązanie tych dwóch zbiorów, wykorzystując kolumnę daty. W wielu przypad-

kach data jest zapisywana na różny sposób. Przykładowo, w jednym zbiorze data jest zapisywana w formacie RRRR-MM-DD, a w drugim RRRR/MM/DD. Należy jednak mieć na uwadze, że wykorzystując narzędzie MS Excel, istnieje możliwość obróbki danych na etapie ich importowania. Dlatego nie powinno to sprawić większego problemu nawet mało doświadczonemu analitykowi czy osobie niemającej technicznej wiedzy informatycznej.

W opisywanym przypadku okazało się, że sposób zapisu danych w archiwizowanych zbiorach z walutami oraz indeksami giełdowymi jest taki sam i ma następujący wzorzec: RRRR-MM-DD. Tym samym możliwe stało się połączenie tych dwóch zbiorów danych bez konieczności zmiany formatu zapisu daty w celu jej ujednoczenia. Jeżeli byłaby taka konieczność, pokazane tabele można byłoby transponować, a poszczególne kursy również utworzyć w postaci własnych klasyfikacji. Miałoby to np. zastosowanie jeżeli łączonych byłoby wiele zbiorów giełdowych.

Kolejnym etapem było powiązanie tych dwóch zbiorów danych. Umożliwiło to zapisanie wszystkich danych w ramach jednej tabeli lub wykresu. Na rysunku 7.3 przedstawiono logiczne złączenie dwóch zbiorów danych, które zostało wykorzystane w opisywanym przypadku.

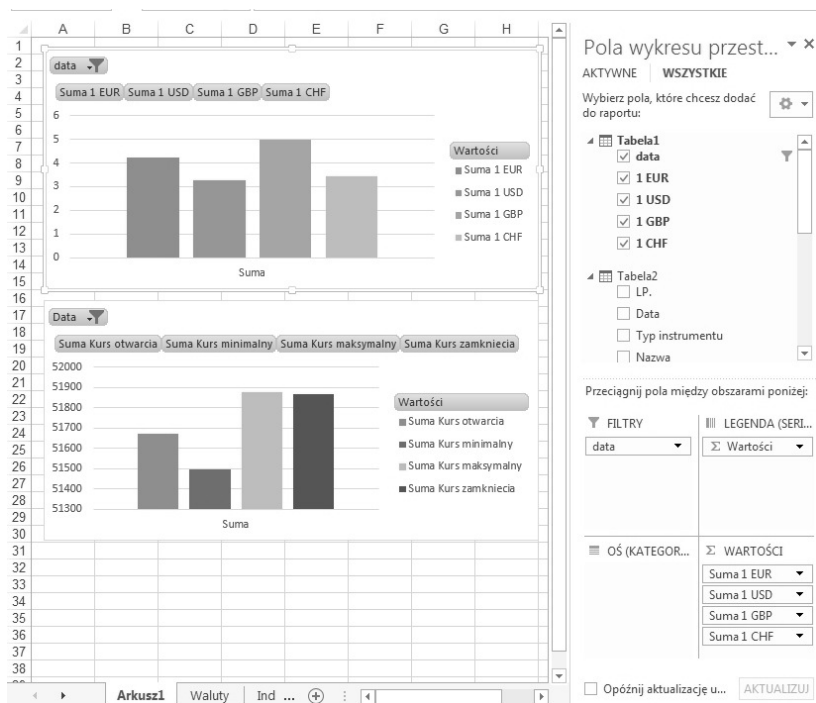


Rysunek 7.3. Logiczne złączenie tabel na potrzeby opisywanego studium przypadku

Źródło: Opracowanie własne.

Przedstawione na rysunku 7.3 pola wskazują, że jako klucze zostały wykorzystane kolumny o nazwie *Data*. W tym przypadku nie ma jednak konieczności łączenia dwóch tabel w jedną tabelę, ze względu na wymagania, jakie firma postawiła przed osobami wdrażającymi kokpit. Zgodnie z założeniami przekazanymi do firmy kluczowe było prezentowanie tych danych na osobnych wykresach lub tabelach. Tym samym zbiory mogły istnieć niezależnie. Jednak powiązanie ich przynosi korzyści w postaci upewnienia się, że szereg czasowy jest zbieżny i można nim sterować za pomocą jednego pola daty.

Zgodnie jednak z wymogami firmy istotne było utworzenie kokpitu menedżerskiego, który zaprezentuje wymagane dane na osobnych wykresach. W tym celu zastosowano typowe narzędzie do prezentacji tabel i wykresów przestawnych, wchodzące w skład dodatku Power Pivot. Zostało to zaprezentowane na rysunku 7.4.



Rysunek 7.4. Kokpit interaktywny z wykorzystaniem wykresów przestawnych

Źródło: Opracowanie własne.

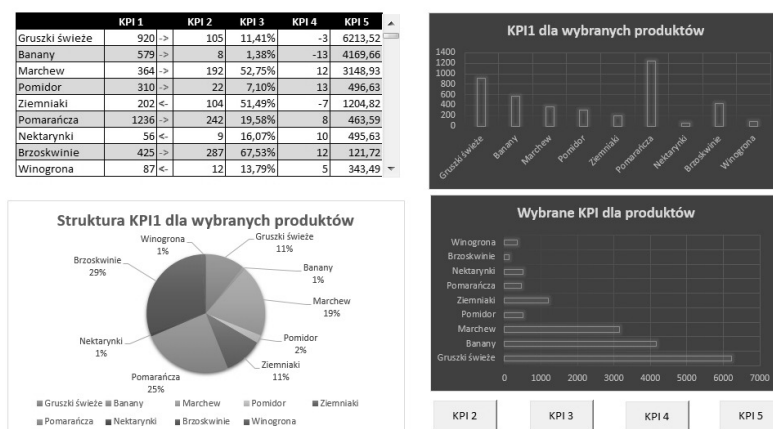
W wyniku zastosowania mechanizmu Power Pivot utworzony został kokpit menedżerski (plik *r7_kokpit_1.xlsx*), który według klasyfikacji użytej w niniejszym podrozdziale zaliczany może być do poziomu 3 kokpitów, czyli zintegrowana kontrola oraz charakter analityczny kokpitu. Ten charakter analityczny objawia się przede wszystkim występowaniem szeregu różnych funkcji statystycznych i matematycznych do dyspozycji analityka. Są one dostępne w oknie *Wartości* w ramach paska bocznego prezentującego pola wykresu przestawnego. Dzięki temu analityk ma możliwość zmiany układu wykresów oraz danych na nich prezentowanych. Umożliwia to elastyczne dopasowanie prezentowanych informacji do aktualnych potrzeb informacyjnych Firmy X.

7.2. Przypadek 2 – kokpit menedżerski z wykorzystaniem KPI do analizy i oceny sprzedaży

Firma X zgłosiła zapotrzebowanie na wdrożenie systemu do monitorowania sprzedaży oferowanych produktów. W tym celu zdecydowano się przygotować podstawowy zestaw KPI dla określania aktualnego stanu sprzedaży produktów. Określono pięć podstawowych wskaźników KPI:

- KPI 1: wielkość sprzedaży ogółem w ostatnim roku działalności Firmy X,
- KPI 2: wielkość sprzedaży w porównaniu z miesiącem ubiegłym,
- KPI 3: wielkość sprzedaży w stosunku do poprzedniego dnia,
- KPI 4: procent przychodów,
- KPI 5: zysk ze sprzedaży produktów.

Na rysunku 7.5 zaprezentowano przykład kokpitu menedżerskiego obejmującego wymienione powyżej wskaźniki KPI.



Rysunek 7.5. Kokpit interaktywny prezentujący wybrane pięć KPI

Źródło: Opracowanie własne.

Przedstawiony na rysunku 7.5 kokpit menedżerski został podzielony na cztery moduły:

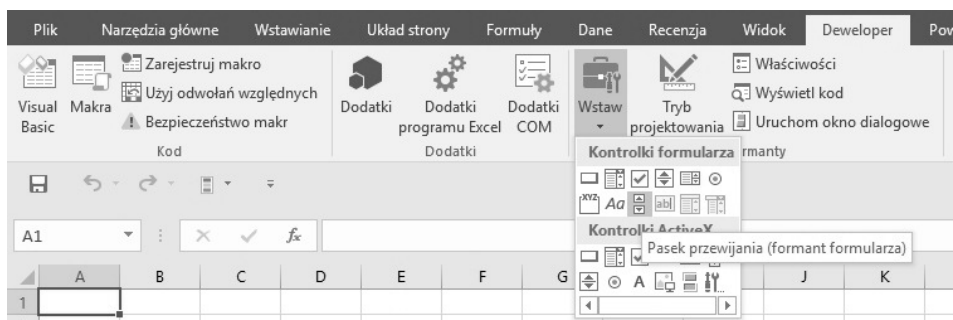
- KPI w formie tabelarycznej z paskiem przewijania,
- wykres kolumnowy prezentujący wartość KPI 1 dla aktualnie wybranych produktów,
- wykres kołowy prezentujący strukturę produktów dla KPI 1,
- wykres słupkowy z możliwością wyboru KPI (2, 3, 4 lub 5).

W celu realizacji pierwszego modułu przyjęto, że pasek przewijania będzie obsługiwany za pomocą formuł MS Excel pozwalających na prezentowanie wskazanych danych pochodzących z dużego zbioru danych. W tym celu przygotowano

zbiór danych w postaci kolumnowej, tj. obliczono wartości KPI dla ostatnich 12 miesięcy. Odpowiednio KPI prezentuje następujące wartości:

- KPI 1 – suma sprzedanych produktów w ostatnich 30 dniach,
- KPI 2 – wzrost/spadek ilości sprzedawanych produktów w stosunku do miesiąca poprzedniego,
- KPI 3 – procentowy wzrost/spadek w stosunku do poprzedniego dnia,
- KPI 4 – wzrost/spadek w stosunku do tygodnia poprzedniego,
- KPI 5 – zysk/strata na sprzedaży produktu w ostatnim miesiącu.

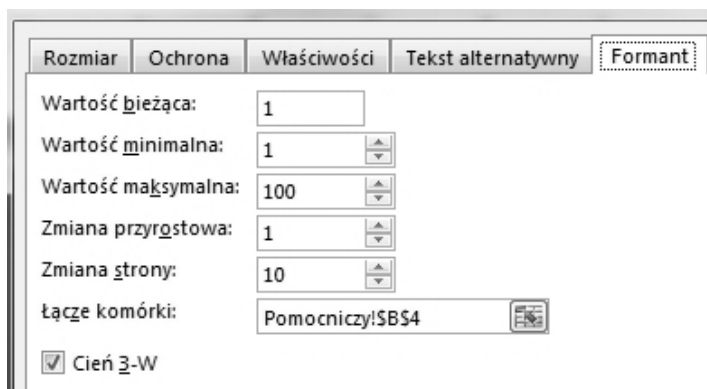
Aby prezentować dane w postaci tabeli pozwalającej na zmianę prezentowanych wartości za pomocą paska przewijania, zdecydowano się dodać formant o nazwie *Pasek przewijania*. Formant ten znajduje się w grupie *Kontrolki formularza* na wstążce *Deweloper* (rysunek 7.6).



Rysunek 7.6. Pasek przewijania dostępny w ramach kontrolki formularza

Źródło: Opracowanie własne.

Dostępne w menu kontekstowym właściwości paska przewijania zostały ustawione na wartości zgodne z rysunkiem 7.7.



Rysunek 7.7. Zmienne ustawione w ramach paska przewijania

Źródło: Opracowanie własne.

Jak widać na rysunku 7.7, minimalna wartość prezentowana na formularzu to 1, maksymalna to 100. Zmiana przyrostowa następuje o 1 (kliknięcie na strzałce paska przewijania) lub o 10 (kliknięcie na polu poniżej lub powyżej aktualnej pozycji).

Aby dane zostały właściwie zaprezentowane w korespondującej do paska tabeli, wykorzystano formułę o nazwie PRZESUNIĘCIE. Zgodnie z dokumentacją firmy Microsoft jej definicja jest następująca:

PRZESUNIĘCIE(odwołanie;wiersze;kolumny;[wysokość];[szerokość])

gdzie:

- odwołanie – określa komórkę lub zakres sąsiadujących komórek stanowiących źródło danych;
- wiersze – liczba wierszy w górę lub w dół, o które należy przesunąć lewą górną komórkę; może być dodatni (przesunięcie w dół) lub ujemny (przesunięcie w górę);
- kolumny – liczba kolumn w lewo lub w prawo, o które należy przesunąć lewą górną komórkę wynikową; może być dodatni (przesunięcie w prawo) lub ujemny (przesunięcie w lewo);
- wysokość – argument opcjonalny – liczba wierszy, którą ma mieć zwracane odwołanie;
- szerokość – argument opcjonalny – liczba kolumn, którą ma mieć zwracane odwołanie [Microsoft, 2015].

W opisywanym przypadku wykorzystano następujący zapis:

=PRZESUNIĘCIE(Dane!C2;Pomocniczy!\$B\$4;0)

gdzie arkusz *Dane* zawiera dane źródłowe z wyliczonymi wskaźnikami KPI, natomiast *Pomocniczy* wskazuje na komórkę (B4), która przechowuje aktualną wartość pozycji paska przewijania. Podobna formuła została zastosowana dla pozostałych KPI.

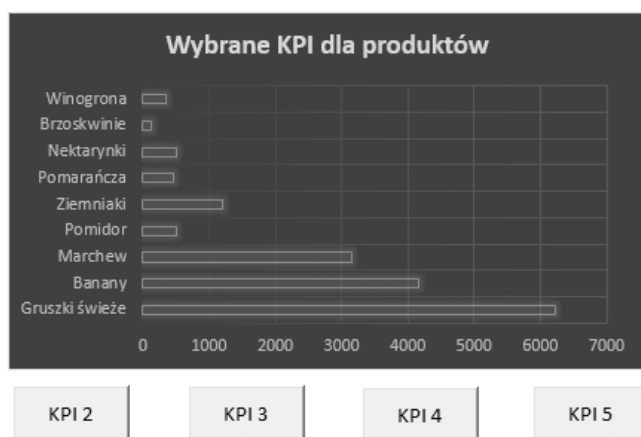
Należy zwrócić uwagę, że zastosowano również dodatkową formułę pozwalającą na wyliczanie ilości sprzedawanych produktów. Jeżeli ilość jest mniejsza niż 0,2 maksymalnej ilości dla wszystkich transakcji, wówczas wartość ta jest oznaczana na czerwono. W innym przypadku na zielono. W celu realizacji tego wymagania wykorzystano następującą formułę:

=JEŻELI(C4/MAX(Dane!C2:C101)<0,2;"<-";
JEŻELI(C4/MAX(Dane!C2:C101)>=0,2;"->"))

Szerzej na temat budowy formuły *JEŻELI* można przeczytać w rozdziale 3. Oznacza to, że znakiem „->” wyróżniane będą te produkty, które bardzo obciążają zasoby magazynowe. Jeżeli produkty będą sprzedawane w niewielkich ilościach (<20% przeciętnej ilości dla transakcji produktów), wówczas będą oznaczane znakiem „<-”. Jednocześnie zastosowano *Formatowanie warunkowe* w celu wyróżniania tych wartości kolorem.

Wykresy będące modułami 2 i 3 są budowane na bieżąco, w momencie wystąpienia zmiany na pasku przewijania. W celu ich budowy wykorzystano typowy mechanizm tworzenia wykresów opisany w rozdziale 4.

Istotnym mechanizmem z punktu widzenia automatyzacji pracy w kokpicie menedżerskim jest wykorzystanie *Przycisków poleceń* do prezentacji danych aktualnie wymaganych. Tak został przygotowany moduł 4, czyli wykres prezentujący KPI 2, 3, 4 lub 5, w zależności od wciśniętego przycisku. Zostało to zobrażone na rysunku 7.8.



Rysunek 7.8. Wybór KPI przełączany przyciskami

Źródło: Opracowanie własne.

W celu prawidłowego wyświetlania wartości wykorzystano fragment kodu VBA (*Visual Basic for Applications*), pozwalający na przenoszenie wartości do tymczasowego arkusza z danymi dla wybranego rodzaju KPI:

```
Dim i As Integer
For i = 4 To 12
    Sheets(3).Cells(i + 10, 4).Value = Sheets(1).Cells(i, 2).Value
    Sheets(3).Cells(i + 10, 5).Value = Sheets(1).Cells(i, 6).Value
Next
```

Zaprezentowane w ramach niniejszego studium przypadku rozwiązanie (plik *r7_kokpit_2*) jest bardzo przyjazne ze względu na możliwość interaktywnego przełączania pomiędzy różnego rodzaju produktami oraz wybranymi KPI. Ma jednak również wiele wad. Pierwsza to niedostępność narzędzi analitycznych takich jak w tabelach przestawnych. Analityk może dane sumować i liczyć średnią za pomocą standardowych narzędzi MS Excel. Rozwiązaniem mogłoby być dodanie kolejnych przycisków, które odpowiednio liczyłyby średnią, medianę czy inne wymagane statystyki.

Kolejną wadę stanowi brak elastycznego wyboru produktów. Analityk może wybrać grupę produktów uporządkowanych według zawartego we wzorcu uporządkowania – np. w kolejności alfabetycznej lub następujących po sobie dat. Zatem kluczowe jest właściwe posortowanie zbioru danych zanim zostanie on wczytany przez narzędzia. Rozwiązanie to mogłoby zostać usprawnione poprzez dodanie opcji wyboru określonego zestawu produktów.

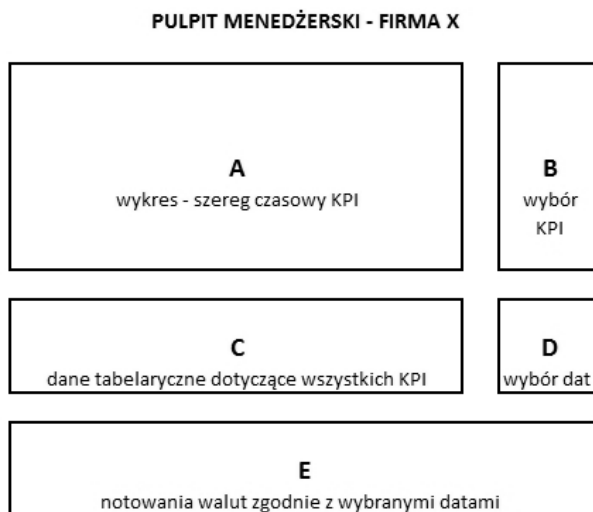
7.3. Przypadek 3 – integracja wielu źródeł danych w jednym kokpicie menedżerskim

Jak wspomniano we wstępie do niniejszego rozdziału, istotne wydaje się przekazanie jak największej ilości informacji w możliwie syntetycznym ujęciu. Zwykle menedżerowie nie dysponują wystarczającą ilością czasu, aby przeglądać rozbudowane kokpity menedżerskie. Wyzwaniem okazuje się zatem wyselekcjonowanie informacji, które są niezbędne w procesie podejmowania decyzji. W warunkach dynamicznie zmieniającego się otoczenia, i tym samym kryteriów mających wpływ na podejmowanie decyzji, niezbędne wydaje się ciągłe analizowanie potrzeb decydentów i dostosowywanie kokpitów menedżerskich do aktualnych potrzeb. Zmiany takie nie powinny jednak zachodzić częściej niż raz w roku.

W niniejszym podrozdziale opisano studium przypadku Firmy X, w której zdecydowano, że menedżerowie będą mieli zamieszczone kluczowe informacje dotyczące sprzedaży produktów (KPI wymienione w podrozdziale 7.2. oraz aktualnie panującej sytuacji na rynku walutowym (podrozdział 7.1.). W związku z powyższym przygotowano dwa źródła danych:

- dane o sprzedaży w Firmie X,
- dane pochodzące ze strony NBP nt. notowań walut.

Dane te zostały przekształcone, aby możliwe było ich zastosowanie w późniejszej analizie. Zdecydowano się na rozmieszczenie komponentów kokpitu menedżerskiego zgodnie z rysunkiem 7.9.



Rysunek 7.9. Projekt funkcjonalny zawartości kokpitu menedżerskiego dla firmy X

Źródło: Opracowanie własne.

Zaprezentowany na rysunku 7.9 schemat posiada następujące odzwierciedlenie w źródłach danych:

- systemy transakcyjne ERP i CRM: komponent A oraz C;
- dane NBP nt. notowań walut: E.

Dodatkowo dane w blokach A, C oraz E są zależne od paska przewijania D (wybór dat). Natomiast blok A jest dodatkowo zależny od wybranego przycisku na komponencie B.

W celu implementacji tego rozwiązania wykorzystano podstawowe mechanizmy dostępne w MS Excel, tj.:

- przyciski sterowane kodem VBA (opis w podrozdziale 7.3),
- pasek przewijania (opis w podrozdziale 7.3),
- standardowe wykresy (zgodnie z opisem w rozdziale 4),
- tabele z formatowaniem warunkowym (opis w rozdziale 5).

Dane stanowiące źródło dla komponentów A, C i E zostały zamieszczone w osobnych dwóch arkuszach (*sprzedaz.xlsx* oraz *kursy_walut.xlsx*). Sposób ich importowania jest zgodny z opisem w rozdziale 1. Mając na uwadze wysoką jakość danych, jak wspomniano na początku podrozdziału, konieczne było przetworzenie ich zgodnie z instrukcjami zawartymi w rozdziale 2. W zależności od rodzaju źródła danych problemy z nieoczyszczonymi danymi mogą doprowadzić do błędnych analiz i – w rezultacie – do niepoprawnych interpretacji, co może skutkować podejmowaniem niewłaściwych decyzji.

7.4. Możliwe zastosowania MS Excel w biznesie w formie zadań

W celu zobrazowania możliwości zastosowania narzędzia Power Pivot w programie MS Excel można posłużyć się poniższymi zadaniami:

1. Firma X chciałaby sprawdzić, na ile sytuacja na giełdzie ma wpływ na notowania walut. Należy sprawdzić korelację pomiędzy wartością polskiej waluty w stosunku do innych walut zagranicznych a wielkością punktów w indeksach giełdowych. W tym celu należy pobrać zbiory danych dotyczące walut oraz danych giełdowych, zgodnie z informacjami znajdującymi się w rozdziale 1 niniejszego opracowania, a następnie zintegrować je zgodnie z informacjami w niniejszym rozdziale.

2. Firma X chce przygotować rozwiązanie bazujące na wskaźnikach KPI badających sprawność organizacji. W tym celu należy zaproponować wskaźnik KPI, który w sposób całościowy obejmie wszystkie sprzedawane produkty i będzie porównywał ich sprzedaż dzień po dniu. W takiej sytuacji nie można agregować danych w tabeli na poziomie produktów, a jedynie na poziomie dat.

3. Firma X chciałaby uzyskać dane na temat możliwości otwarcia oddziału w jednym z wybranych krajów Unii Europejskiej. W związku z tym należy utworzyć kokpit menedżerski, który pokaże sytuację związaną z bezrobociem i wykształceniem osób w zależności od wybranego kraju, jednocześnie porównując dane ze średnią dla wybranych krajów. Przełączanie pomiędzy krajami może następować za pomocą paska przewijania lub przycisków.

4. W Firmie X planuje się dodanie kolejnego wskaźnika KPI, tym razem monitorującego, jaka jest sezonowa sprzedaż produktów. W związku z tym należy wyliczyć nowy zbiór danych agregujący informacje z tym związane według kwartałów. Następnie wartości te powinny zostać zaprezentowane na wykresie, uwzględniając wskaźnik sprzedaży w bieżącym kwartale w stosunku do kwartału poprzedniego, oznaczając go jako kolejny KPI.

5. Firma X planuje podjąć działania związane z integracją danych opisywanych w ramach zastosowań wymienionych we wcześniejszych punktach. Należy zaproponować i wdrożyć rozwiązanie integrujące utworzone wcześniej moduły.

Zakończenie

Analiza możliwości zastosowania arkusza MS Excel jednoznacznie wskazuje, że istnieje duży niewykorzystany potencjał tego narzędzia również w zakresie analityki biznesowej.

Istotne jest jednak, aby tworzone oprogramowanie spełniało cechy wymienione we wstępie do niniejszego rozdziału, tj. intuicyjność, interaktywność, charakter

analityczny i przejrzystość. Tworząc tego rodzaju rozwiązania, należy sięgnąć po uznane przez analityków narzędzia, patrząc jednak z perspektywy wszechstronnego zastosowania narzędzia. Wydaje się, że narzędzia dedykowane, przystosowane tylko do pełnienia funkcji kokpitu menedżerskiego, mogą być niezbyt właściwe dla tego rodzaju zastosowań.

Dlatego w opisywanych studiach przypadków zdecydowano się na wykorzystanie powszechnie znanego narzędzia, jakim jest MS Excel. Pozwoliło to na utworzenie sprawnego narzędzia, posiadającego požądane przez analityków cechy interaktywności, analityki biznesowej i przejrzystości interfejsu użytkownika.

Bibliografia

1. Bradea I., Sabău-Popa D., Boloş M. (2014), *Using Dashboards in Business Analysis*, Annals Of The University Of Oradea, Economic Science Series, Vol. 23, No. 1.
2. Bremser W., Wagner W. (2013), *Developing Dashboards for Performance Management*, w: „CPA Journal”, Vol. 83, No. 7.
3. Guni C. (2014), *The Dashboard – Conceptual Dimensions and Evolutions*, „Economics, Management & Financial Markets”, Vol. 9, No. 1.
4. Holocaine D., Airinei D. (2015), *On-Line Dynamic Dashboards in Audit Activities*, „Audit Financiar”, Vol. 13, Issue 125.
5. Hsinchun C., Chiang R., Storey V. (2012), *Business Intelligence and Analytics: From Big Data do Big Impact*, „MIS Quarterly”, Vol. 36, No. 4.
6. Lautman M., Pauwels K. (2013), *What Are the Real Key Performance Indicators (KPIs) That Drive Consumer Behavior?*, „GfK-Marketing Intelligence Review”, Vol. 5, No. 2.
7. Manzetti P., Mehta A. (2015), *How to Engage and Inform the Nonprofit Board The Value of the Performance Dashboard*, „CPA Journal”, Vol. 85, No. 6.
8. Sangeeta S., Malarvizhi V. (2015), *Dashboard to show key metrics for teams following agile methodology*, „International Journal of Applied Engineering Research”, Vol. 10, No. 12.
9. Shadpour F., Kilcoyne J. (2015), *Criteria for Building Automation Dashboards*, „Ashrae Journal”, Vol. 57, No. 5.
10. Wrycza S. (red.) (2010), *Informatyka Ekonomiczna. Podręcznik akademicki*, PWE, Warszawa.
11. <https://support.office.com/pl-pl/article/PRZESUNI%C4%98CIE-funkcja-c8de19ae-dd79-4b9b-a14e-b4d906d11b66>, dostęp dnia: 15.09.2015 r.

Wykorzystane skróty

CN – *combined nomenclature*

CRM – *Customer Relationship Management* – zarządzanie relacjami z klientami

CSV – *Comma Separated Value*

DAX – *Data Analysis Expressions*

DDDM – *Data Driven Decision Making*

ERP – *Enterprise Resource Planning* – planowanie zasobów przedsiębiorstwa

KPI – *Key Performance Indicator* – kluczowy miernik wydajności

MS – Microsoft

ODBC – *Open DataBase Connectivity*

OLAP – *On-Line Analytical Processing*

SQL – *Structured Query Language*

TCP/IP – *Transmission Control Protocol/Internet Protocol*

VBA – *Visual Basic for Applications*

Pliki z danymi oraz wynikami analizy danych

Pliki zawierające dane wejściowe i wyniki analiz wykorzystane w ramach poszczególnych rozdziałów i studiów przypadków, dostępne pod adresem <https://sites.google.com/site/kokpitymenedzerskiewanalizie/>:

Nr rozdziału	Studium przypadku	Dane wejściowe (nazwa pliku, arkusza)	Dane wyjściowe (nazwa pliku, arkusza)
Rozdział 1	Przypadek 1	-	r1_import_i_eksport_no.xlsx
	Przypadek 2	-	r1_dane_sprzedazy_no.csv
	Przypadek 3	-	r1_kursy_walut_2015.csv; r1_kursy_walut.xlsx
	Przypadek 4	-	r1_indeks_WIG.xls; r1_dane_gieldowe.xlsx
	Przypadek 5	-	r1_wyzsze_wyksztalzenie.xlsx; r1_wynagrodzenia_brutto.xlsx
Rozdział 2	Przypadek 1	r1_import_i_eksport_no.xlsx	r2_import_i_eksport_o.xlsx
	Przypadek 2	r1_dane_sprzedazy_no.csv	r2_dane_sprzedazy_o.xlsx
	Przypadek 3	r1_faktury_sprzedazy_no.xlsx	r2_faktury_sprzedazy_o.xlsx
Rozdział 3	Przypadek 1	r3_analiza_sprzedazy_formuly.xlsx arkusz dane sprzedaży	r3_analiza_sprzedazy_formuly.xlsx arkusz analiza sprzedaży_1
	Przypadek 2	r3_analiza_sprzedazy_formuly.xlsx arkusz analiza sprzedaży_1	r3_analiza_sprzedazy_formuly.xlsx arkusz analiza sprzedaży_2
	Przypadek 3	r1_wyzsze_wyksztalzenie.xlsx	r3_wyzsze_wyksztalzenie_formuly.xlsx
	Przypadek 4	r3_analiza_sprzedazy_formuly.xlsx arkusz analiza sprzedaży_2	r3_analiza_sprzedazy_formuly.xlsx arkusz analiza sprzedaży_4
	Przypadek 5	r3_analiza_sprzedazy_formuly.xlsx arkusz analiza sprzedaży_4	r3_analiza_sprzedazy_formuly.xlsx arkusz analiza sprzedaży_5
Rozdział 4	Przypadek 1	r4_wynagrodzenia_brutto.xlsx arkusz wynagrodzenia	r4_wynagrodzenia_brutto.xlsx arkusz wynagrodzenia_form warun
	Przypadek 2	r4_wyzsze_wyksztalzenie.xlsx arkusz wykształcenie	r4_wyzsze_wyksztalzenie.xlsx arkusz wykształcenie_wykr przeb
	Przypadek 3	r4_kursy_walut.xlsx arkusz kursy	r4_kursy_walut.xlsx arkusz kursy_wykres
	Przypadek 4	r4_faktury_sprzedazy_xlsx, arkusz zysk_strata arkusz formy płatności	r4_faktury_sprzedazy_xlsx, arkusz zysk_strata_wykres arkusz formy płatności_wykres
	Przypadek 5	r4_wyzsze_wyksztalzenie.xlsx arkusz wykształcenie	r4_wyzsze_wyksztalzenie.xlsx arkusz wykształcenie_mapa

Rozdział 5	Przypadek 1	r5_faktury_sprzedazy_sumycz.xlsx arkusz faktury sprzedaży	r5_faktury_sprzedazy_sumycz.xlsx arkusz sumy częściowe_1 arkusz sumy częściowe_2
	Przypadek 2	r5_faktury_sprzedazy_filtr.xlsx, arkusz faktury sprzedaży	r5_faktury_sprzedazy_filtr.xlsx, arkusz fragmentatory arkusz czyszczenie
	Przypadek 3	r5_faktury_sprzedazy_filtr.xlsx, arkusz faktury sprzedaży	r5_faktury_sprzedazy_filtr.xlsx, arkusz faktury sprzedaży
	Przypadek 4	r5_import_i_eksport_filtr.xlsx arkusz EKSPORT_2014 arkusz Kraje i kontynenty	r5_import_i_eksport_filtr.xlsx arkusz EKSPORT_2014_o arkusz Kategorie krajów arkusz eksport_lista krajów
	Przypadek 4	r5_faktury_sprzedazy_tabele.xlsx, arkusz faktury sprzedaży	r5_faktury_sprzedazy_tabele.xlsx, arkusz analiza wg produktów arkusz analiza wg klientów arkusz analiza wg sprzedawców arkusz analiza wg obszarów arkusz analiza w czasie
	Przypadek 5	r5_faktury_sprzedazy_tabele.xlsx, arkusz faktury sprzedaży	r5_faktury_sprzedazy_tabele.xlsx, arkusz analiza w czasie
Rozdział 6	Przypadek 1	r1_kursy_walut_2015.csv ; r1_kursy_walut.xlsx r1_dane_gieldowe.xlsx r2_dane_sprzedazy_o.xlsx	r6_sprzedaz_gielda_waluty_model .xlsx
	Przypadek 2	r2_faktury_sprzedazy_o.xlsx	r6_faktury_sprzedazy_model.xlsx
	Przypadek 3	r6_faktury_sprzedazy_model.xlsx	r6_faktury_sprzedazy_DAX.xlsx
	Przypadek 4	r2_import_i_eksport_o.xlsx	r6_import_i_eksport_model.xlsx
Rozdział 7	Przypadek 1	r1_kursy_walut.xlsx r1_dane_gieldowe.xlsx	r7_kokpit_1.xlsx
	Przypadek 2	r2_dane_sprzedazy_o.xlsx	r7_kokpit_2.xlsx
	Przypadek 3	r2_dane_sprzedazy_o.xlsx	r7_kokpit_3.xlsx

Nota o autorach

Dorota Buchnowska, dr, Katedra Informatyki Ekonomicznej, Wydział Zarządzania, Uniwersytet Gdański, dorota.buchnowska@ug.edu.pl

Dariusz Krlewski, dr, Katedra Informatyki Ekonomicznej, Wydział Zarządzania, Uniwersytet Gdański, darek@ug.edu.pl

Michał Kuciapski, dr, Katedra Informatyki Ekonomicznej, Wydział Zarządzania, Uniwersytet Gdański, m.kuciapski@ug.edu.pl

Bartosz Marcinkowski, dr, Katedra Informatyki Ekonomicznej, Wydział Zarządzania, Uniwersytet Gdański, bmarc@ug.edu.pl

Jacek Maślankowski, dr, Katedra Informatyki Ekonomicznej, Wydział Zarządzania, Uniwersytet Gdański, jacek.maslankowski@ug.edu.pl

Monika Woźniak, dr, Katedra Informatyki Ekonomicznej, Wydział Zarządzania, Uniwersytet Gdański, monika@ug.edu.pl